Resource allocation and risk assessment in pandemic situations

D i s s e r t a t i o n

zur Erlangung des akademischen Grad des
Doctor rerum naturalium
(Dr. rer. nat.)

eingereicht an der
Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin
von

Olga Baranov

Präsidentin/Präsident der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekanin/Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin
Prof. Dr. Bernhard Grimm

Gutachter/innen

1. Prof. Dr. Dirk Brockmann

2. Prof. Dr. Vitaly Belik

3. Dr. habil. Philipp Hövel

Tag der mündlichen Prüfung: 12.10.2018

I hereby declare that I completed the doctoral thesis independently based on the stated resources and aids.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und nur mit den angegebenen Hilfsmitteln erstellt habe

Olga Baranov
5th January 2019

Abstract

The growing complexity of the global mobility is a key challenge for the understanding of the worldwide spread of emergent infectious diseases and the design of effective containment strategies. Despite global connectivity, containment policies are often based on national, regional and 'egocentric' assessments of outbreak situations that are no longer effective or meaningful, as recently demonstrated by 2014 Ebola outbreak in West Africa, where months passed before a concerted, international effort followed. Despite the importance of the matter, optimal strategies in highly connected non-local settings are poorly understood.

In the work at hand we propose a set of methods for more informed decision making during and prior to a pandemic. We introduce a method to calculate the risk of disease importation in a specific location, propose metrics which characterise the role of a node during a distinct outbreak and investigate the allocation of resources as a game theoretic dilemma. All of the studied systems are represented by networks of connected nodes to account for the dominant role played by the traffic networks during the global spreading of the disease. While multiple publications embraced a minimalistic approach by concentrating on network features and disregarding the specifics of the outbreak, others employed elaborate simulations which are highly reliant on a plethora of disease specific parameters. To foster the benefits of both approaches, the work at hand will follow the intermediate path by concentrating on the network topology with regard to the origin of the outbreak. Presented work relies on the concept of effective distance introduced by Brockmann and Helbing in [17].

First, we devise two metrics which characterise the role played by a node during a pandemic scenario. Using the effective distance trees we define the *scope* of a node, which shows the fraction of the network population that benefits from the countermeasures deployed at the respective node. Since a travelling infected has to traverse a node in question to reach the downstream population, the exit screening at the node in question will have a protective effect for the downstream population. The second metric, *confluence*, characterises the branching of the effective distance tree after the node in question. This metric highlights the nodes which are located at the end of a bottleneck on the effective distance tree. They are a good fit for countermeasure deployment as a high amount of paths are blocked by successful screening and containment at such a node. In combination, both metrics can be used to narrow down the set of candidate nodes for countermeasure deployment and be used for final decision making on par with other determining factors. We demonstrate the usage of the metrics on real-world and hypothetical scenarios.

Before approaching the resource allocation problem, we device a method to estimate the *import risk* of disease into a node, as it is necessary to judge the threat posed by a specific outbreak. To account for the specific behaviour of agents on transportation networks, we define the probability that an agent will exit the traffic network at a specific node when starting at a defined location. The *exit probability* is not uniform across nodes as the traffic connections between two remote locations often involve the transit via multiple additional hubs. While this information can be extracted from data when available, we show how this probability can be derived from the structure of the effective distance tree. Using the exit probability and the information about the origin of the outbreak, we calculate the probability of case importation into a node and hence the threat posed by an outbreak. Using real-world and hypothetical outbreak scenarios as examples we demonstrate how the import risk can be averaged over a broad region to reflect the spread of the pandemic.

Finally, we investigate the resource allocation during a pandemic scenario from selfish and pro-social perspective. The question is approached using the cost function optimisation with the function varying depending on the mode of optimisation. The problem resembles the public good dilemma well known in the game theory but with an important difference:

While the basic form of the cost function is equal for every node, the import risk varies due to different positions on the network. We show that in case of an outbreak concentrated in a single location the selfish and pro-social outcomes do not differ substantially. The allocation of resources to the source of the outbreak is the optimal strategy in both cases. When the pandemic is seeded at two remote locations, self-investment emerges as an optimal strategy for a subset of nodes in the selfish scenario. Furthermore, we demonstrate that the risk of case importation and the ratio of import risks from multiple locations are the determining factors when it comes to the resource allocation decisions. Multiple additional insights can be gained from the simplified model we are presenting, e.g. that self-investment is never pro-socially optimal, and a strategy resembling the ring vaccination emerges when investment into the infected is not possible.

Zusammenfassung

Das Verständnis der komplexen Interaktionen innerhalb des weltweiten Transportnetzes ist ein essentieller Schritt auf dem Weg zur Vorhersage der Krankheitsausbreitung und Entwicklung von effektiven Gegenmaßnahmen. Ungeachtet der weltweiten Vernetzung werden die politischen Entscheidungen oft von nationaler, regionaler und egozentrischen Denkweise geleitet. Die Ebola-Epidemie in 2014 demonstrierte deutlich, dass solche Herangehensweise modernen Epidemien nicht gerecht werden kann. Trotz des Potentials viele Menschenleben zu retten, ist die optimale Strategie im Falle eines Ausbruchs in einer vernetzten, komplexen Umgebung nicht ausreichend verstanden.

In dieser Dissertation werden mehrere Methoden entwickelt, welche es ermöglichen während einer Epidemie die globalen Teilnehmer entsprechend ihrer Rolle einzustufen und das Risiko des Krankheitsexports zu berechnen. Darüber hinaus wird mit Hilfe eines vereinfachten spieltheoretischen Modells der Prozess der Ressourcenverteilung zur Epidemieeindämmung untersucht. In allen Aspekten der Arbeit wird netzwerkbasierte Repräsentation des Systems verwendet, so dass die Orte durch knoten innerhalb eines Transportnetzwerkes abgebildet werden. Viele Publikationen, die die Eindämmungsstrategien im Kontext von Epidemien untersucht haben, konzentrierten sich auf der Netzwerktopologie und ließen die krankheitsspezifischen Parameter außer Acht. Wiederum andere nutzten aufwändige Simulationen, welche entscheidend von der Wahl der epidemiologischen Parameter abhängen. Um die Vorteile beider Herangehensweisen auszunutzen wird in dieser Arbeit der Mittelweg eingeschlagen: während der Schwerpunkt auf der Topologie des Netzwerkes liegt, berücksichtigt die vorgestellte Methodik den Ursprungsort der Epidemie. Die vorgestellte Arbeit nutzt das Konzept der effektiven Distanzen die darauf basierenden kürzesten Pfade und Bäume, vorgestellt in [17].

Um die Rolle eines Knoten im Kontext eines bestimmten Ausbruchs zu charakterisieren definieren wir zwei Zentralitätsmaße: *scope* und *confluence*. *Scope* zeigt an welcher Anteil der Netzwerkpopulation von Eindämmungsmaßnahmen im entsprechenden Knoten profitieren wird. Die zweite Metrik, *confluence*, spiegelt die Verzweigung des Baums der kürzesten Pfade nach dem entsprechenden Knoten wieder. Folglich identifiziert diese Metrik jene Knoten, die für viele wahrscheinlichsten Pfade zwischen dem Ausbruchsort und dem Rest des Netzwerks wichtig sind. Dementsprechend sind Knoten mit hoher *confluence* bevorzugte Ziele für Gegenma{\ss}nahmen. In Kombination können beide Metriken verwendet werden um die besten Kandidaten für den Einsatz der Gegenmaßnahmen auszuwählen. Anhand von realen und hypothetischen Ausbruchsszenarien wird die beschriebene Anwendung demonstriert.

Bevor eine Untersuchung der Ressourcenverteilung gemacht werden kann ist es notwendig eine Methode zu finden die Gefahr abzuschätzen, die von einer Epidemie ausgeht. Diese sind definiert als die Wahrscheinlichkeit, dass Infizierte, die auf der Transportnetzwerk bewegen, die Krankheit in den Knoten eintragen. Um die Spezifik der Transportnetzwerke zu berücksichtigen muss die Wahrscheinlichkeit bestimmt werden, dass ein Individuum, welches sich auf dem Netzwerk bewegt, dieses an einem spezifischen Knoten verlässt. Diese Wahrscheinlichkeit, *exit probability*, kann zwischen den Knoten stark variieren, da Reiseverbindungen zwischen zwei Orten häufig über mehrere Transitknoten erfolgen. Während die *exit probability* aus den Daten entnommen werden kann, sind Informationen darüber nicht immer vorhanden. Daher wird aufgezeigt wie die *exit probability* aus der Struktur des Baumes der kürzesten Pfade geschätzt werden kann. Die Wahrscheinlichkeit des Imports einer Krankheit wird unter Verwendung der *exit probability* und der Information über der Ursprung der Ausbruchs berechnet. Das wird anhand von mehreren realen und hypothetischen Epidemiebeispielen illustriert.

Abschliessend wird ein Modell entwickelt um die Frage der optimalen Ressourcenver-

teilung im Falle einer Pandemie zu studieren. Die Optimalität wird durch die Minimierung einer Kostenfunktion durch einzelne Knoten erreicht. Die Optimierung erfolgt unter einem egoistischen oder pro-sozialem Blickwinkel. Diese Problemstellung ähnelt dem *public good dilemma* aus der Spieltheorie,doch unterscheidet sich in einem entscheidenden Aspekt: die allgemeine Form der Kostenfunktion einzelner Knoten ist gleich, doch der numerische Wert unterscheidet sich dadurch, dass das Risiko des Krankheitsimport in einzelne Knoten stark verschieden sein kann. Das Modell zeigt auf, dass das globale Resultat der Optimierung durch egoistische und pro-soziale Agenten nah bei einander liegt, wenn der Ausbruch in einer einzelnen Quelle beginnt. In beiden Fällen ist es die optimale Lösung die Ressourcen zur Pandemiebekämpfung dem Ausbruchsursprung zu Verfügung zu stellen. Wenn die Epidemie zwei getrennte Quellen aufweist, bildet sich eine neue Strategie heraus, die für egoistische Knoten optimal sein kann. Bei dieser Strategie werden die Ressourcen von dem Knoten selbst für Pandemieprävention utilisiert. Darüber hinaus zeigt das Modell, dass der für die Strategie entscheidende Faktor die Wahrscheinlichkeit des Krankheitsimportes und im Falle von mehreren Ausbruchsquellen der Quotient aus den Wahrscheinlichkeiten ist. Anhand des Modells können weitere Erkenntnisse generiert werden, wie zum Beispiel die Tatsache, dass aus der pro-sozialen Perspektive die Investitionen in Prävention innerhalb von nicht betroffenen Knoten nie optimal sein kann oder dass die Ringimpfung sich als dominante Strategie herauskristallisiert wenn die Investition von Ressourcen innerhalb von dem Ausbruchsort nicht möglich ist.

Contents

1 Introduction **1**

2 State of the Art **11**
2.1 Basic network theory . . . 12
2.2 Path finding, Dijkstra and Yen algorithms . . . 15
2.3 Epidemiology and disease spreading on networks . . . 16
2.4 Game theory and disease containment . . . 18
2.5 Optimisation and evolutionary strategy . . . 20
2.6 Existing approaches for resource allocation . . . 23
2.7 Summary . . . 24

3 Prerequisites **27**
3.1 World air transport network . . . 28
3.2 Effective distance . . . 30

4 Defining the scope: A context specific approach to identifying key airports during a pandemic **33**
4.1 Definition of scope and confluence . . . 35
4.2 Application of the metrics on world aviation network . . . 41
4.3 Discussion . . . 49

5 Quantitative assessment of import risks for pandemic onset situations **51**
5.1 Mathematical definition of import risk . . . 52
5.2 Application on world aviation network . . . 58
5.3 Discussion . . . 69

6 The good, the bad and the optimal: resource allocation strategies during emerging pandemics **71**
6.1 Model . . . 73
6.1.1 Mathematic definition . . . 75
6.1.2 Implementation . . . 81
6.2 Simulated results . . . 83
6.3 Discussion . . . 96

7 Summary and outlook **101**

8 Appendix **111**
8.1 Relation between scope and betweenness centrality . . . 111
8.2 Scope based clusters . . . 111
8.3 Circumstances under which exit probability will not increase along one branch 112
8.4 Taylor expansion of the cost function . . . 112
8.5 Comparison of strategic decisions . . . 113
8.5.1 Invest in n . . . 113

8.5.2 Invest in *a* or invest in *b* . . . 113
8.6 Probability to receive investment . . . 114
8.6.1 Probabilities on a chain . . . 114
8.6.2 Probabilities on a more general topology . . . 115
8.7 Observed investment threshold . . . 116
8.8 Choice of path threshold and filtering algorithm . . . 117
8.8.1 Modified evolutionary strategy algorithm . . . 118
8.9 Tables . . . 118
8.9.1 IATA code of 100 biggest airports . . . 118
8.9.2 List of generalist and specialist airports among 100 biggest on the WAN 121
8.9.3 Default parameter set . . . 124

9 Acknowledgements **127**

References **129**

List of Figures

3.1 Structure of the WAN 29

4.1 Examples of shortest path trees derived from air transportation network 34
4.2 Top 15 airports according to passenger flux , betweenness centrality and closeness centrality. 36
4.3 Graphic explanation of scope and confluence 37
4.4 Relationship between proposed and established centrality measures 39
4.5 Relationship between average scope and confluence. 40
4.6 Scope profiles of big airports with respect to 22 regions of the air transportation network 43
4.7 Hierarchical clustering of top 100 airports according to their regional profiles . 44
4.8 Application of scope on exemplary outbreak scenarios 46
4.9 Relevant outbreak locations from the perspective of a hub airport 48

5.1 Exit probability in an exemplary network 54
5.2 Total import risk over all nodes depending on the maximal path length. . . . 55
5.3 Example of the distribution of import risk (red) and exit probability (blue) in a network 57
5.4 Relationship between import risk , exit probability , scope and betweenness centrality. 58
5.5 Top 20 airports (**A**) and countries (**B**) according to global import risk 59
5.6 Import risk profiles of the biggest airports and respective countries for each continent 61
5.7 Country profiles of China and USA with exemplary airport profiles. 62
5.8 Application of import risk on real and hypothetical outbreak scenarios. 63
5.9 Comparison of scope and import risk profiles and ranking 65
5.10 Import probability in case of a region wide pandemic 66
5.11 Outbreak locations which pose highest threat from the perspective of an airport 68

6.1 Schematic explanation of resource allocation model. 74
6.2 Probability to receive investment with respect to position 80
6.3 Algorithm of game theoretical model implementation. 81
6.4 Implemented susceptibility functions 83
6.5 Resource allocation in a network with one infected 84
6.6 Final state of the model with multiple outbreak sources 86
6.7 pro-social and selfish optimisation with varying amount of resources 89
6.8 Influence of cost of infection and system-wide resources on strategy distribution 91
6.9 Amount of investment into infected in final state of different realisation of Erdős–Rényi networks. 93
6.10 Strategy distribution when strategic restrictions are imposed. 95
6.11 Influence of sources' position on equilibrium state in selfish optimisation case 97

8.1 General network topology used to calculate investment probability. 116
8.2 Path threshold: cost and benefits 117

List of symbols

formula symbol	description
$p_\infty(m\|n_0)$	Import probability from n_0 to m through paths of all possible lengths
F_{mn}	absolute flux from n to m
f_{mn}	relative flux from n to m
P_{mn}	probability that an agent leaving node n will proceed to node m
N	number of nodes in a network
$q_n(n_0)$	exit probability at n given an outbreak at n_0
S_{nm}	probability of an agent to proceed traveling from node m to node n
$\Omega(i,n,t)$	set of all possible paths from node n to node i in t steps
$\Omega_k(i,n,t)$	subset of all possible paths from node n to node i in t steps passing through k
d_{nm}	distance from node m to node n
$T(n_0)$	shortest path tree with n_0 as its origin
N_n	absolute population at node n
η_n	relative population at node n
$\Phi(n\|n_0)$	sum of the absolute flux of node nand its offsprings in the shortest path tree origination at n_0
$\phi(n\|n_0)$	sum of the relative flux of node nand its offsprings in the shortest path tree origination at n_0
D_{mn}	effective distance from node n to node m
$j_0(n)$	infected at node n at time 0
Q_n	containment / susceptibility at n
R_{nm}	investment / resource donation form m into n
W_{ij}	link weight between j and i (equivalent to flux)

formula symbol	description
$C_c(n)$	closeness centrality of n
$C_b(n)$	betweenness centrality of n
S_{ij}	link salience for a link from i to j
T_{ij}^n	risk posed by an outbreak to node n
$c_n(T_i)$	confluence of non an effective distance tree rooted at i
s_n	scope of n
$\alpha(n)$	weighing coefficient of n as one of outbreak origins
λ_n	the number of infected individuals in the source of the outbreak
ρ_n	max. available resources of node n
Γ	relative distribution of the infection force inwards
C_I^0	cost of case importation
C_m	total cost for node m
$\pi_i(t)$	probability to get investment at time step t
Υ	set of all infected in the network
H_n	harmonic number n
r	strategy vector
$\lambda(\omega)$	length of a path ω
$\Theta(m\|n)$	set of nodes which can be accessed via node n given m is the source of the shortest path tree

Table 1: List of formula symbols used throughout the thesis

Chapter 1

Introduction

Since the beginning of the twenty first century multiple large scale outbreaks have demonstrated that in the modern world epidemics can travel at an unprecedented speed and cross large distances. Severe acute respiratory syndrome (SARS) was "the first severe infectious disease to emerge in the twenty-first century" according to the World Health Organisation (WHO) [8]. After surfacing in 2002 it claimed 774 lives and sickened 8096 people over a course of 8 months. Within weeks SARS was present in 26 countries and across 5 continents. Only two years after the SARS pandemic the International Health Regulations (IHR) were signed by all member states of the WHO, underlining that SARS emergence and rapid spread was a wake-up call for the international community [109]. By establishing the IHR, the WHO gained legal grounds to establish binding rules for surveillance and investments in the health infrastructure of countries. Among other tools there exists the declaration of the public emergency of international concern (PHEIC) which enables WHO to call international attention to pandemic events and humanitarian crises posing a threat for the global community. Five years after the SARS epidemic the first PHEIC was declared by the WHO in connection with the H1N1 flu pandemic, an emergent influenza strain of the same subgroup as the Spanish flu [75]. Until today PHEIC was declared four times, with most recent call connected to the increase of microcephaly cases associated with Zika infections. Considering that the IHR entered into force in 2007, it emphasises that global pandemics are a major problem of the twenty-first century, occurring on a regular basis.

Middle east respiratory syndrome (MERS-CoV) was first confirmed as a new zoonotic strain of a corona virus in 2012. Since then 2122 laboratory cases and 740 related deaths were confirmed as of January the 2nd, 2018. While not declared a PHEIC, the disease has raised strong concern and media attention due to its initial similarity to SARS. The MERS-CoV transmission is still ongoing with many transmission events between dromedary camels and humans. Sustained human-to-human transmission was not observed until now. In 2014 Ebola has reached West Africa, resulting in the most severe Ebola outbreak ever recorded and claiming more lives than all outbreaks of this disease since its emergence in 1976 combined [114]. During the time between the onset of the epidemic in January 2014 and the lifting of the PHEIC status in March 2016 Ebola infected 28616 people, killing 11310 of them with mortality rates ranging from 25% to 90% [111, 113]. In the aftermath of the Ebola outbreak, the Ebola Interim Assessment Panel heavily criticised the international response and the delayed reaction of the WHO [77]. Among other points made by the panel was a delayed declaration of the PHEIC, the slow reaction and the lack of financial commitment of the WHO member states or non-adherence to the recommendations of the PHEIC committee.

As demonstrated by multiple epidemics, like those caused by influenza and SARS, the spread of human diseases is tightly coupled to human mobility. Several disease importations of the past centuries occurred in ports by means of the sea travel, dominant long distance

traffic of past ages. Prominent examples are plague, cholera and HIV [101]. The influence of human traffic on the spread of modern epidemics has been observed by multiple studies and is well documented for some pandemics [61, 17, 41]. Considering the increasing availability and speed of long distance traffic, disease in the modern world has the means to spread over huge geographic distances in timespans of hours and days. This has been demonstrated by the SARS epidemic and several influenza outbreaks. Ebola, while not sparking an epidemic beyond the borders of West Africa, has demonstrated how pandemics can unfold when reaching major cities. Upon reaching the capital the case numbers of the epidemic skyrocketed and the containment efforts were heavily complicated. Paired with the growing urbanisation in developing countries causing densities of up to 20000 citizens per kilometre, diseases which were unable to spread in dispersed populations will pose an ever increasing problem in the future. In such conditions contact tracing is hardly feasible and isolation of all possible contacts is impractical. As big cities are also transportation hubs, the importation is likely to occur in one of these, resulting in optimal conditions for a hard to contain epidemic. These challenges were demonstrated in the course of the yellow fever outbreak in 2015 to 2016, where a sustained transmission chain was established in the capital city of Angola, Luanda [112].

Combating a potentially global outbreak is a multilevel process involving a variety of actors. The interruption of local transmission is an indispensable part of the containment effort and it requires high commitment and a considerable amount of resources. Apart from treating infected individuals, medical workers on site have to establish the potential source of infection, the transmission chain and further epidemic parameters. The methodology to solve these questions includes laboratory diagnostics, surveys and questionnaires. Using mathematical modelling projections about the course of the epidemics can be made, but only after disease parameters are known. Two essential disease parameters are the time an infected takes to recover and the number of secondary cases produced by an infected. Establishing the latter can be challenging in the face of an unfolding epidemic. The number of secondary infections is prone to fluctuations, can be varying across different regions and change with time and the underlying social structure. Based on the estimated parameters the final size of the epidemic can be evaluated. Furthermore, the risk of a global pandemic can be judged using these estimates. The mathematical models used for this purpose can range from simple systems of differential equations to sophisticated models including the structure of traffic and behavioural changes in response to the disease. Based on the predictions of models the countermeasures can be devised and implemented by the NGOs and the global community. As has been demonstrated during the Ebola epidemic, a scientific community can contribute a lot to the containment efforts: through theoretical models or by pushing forward the development of a vaccine, better treatment and diagnostic.

The work at hand is attributed to the field of mathematical modelling of pandemic spread and countermeasure deployment. The aim of the broad field is to devise and evaluate computationally the possible intervention strategies to interrupt the local transmission chain or to prevent the outbreak from reaching global dimensions. With computing capacity becoming more accessible, the field of mathematical modelling in the context of epidemic spreading has gained popularity. The use of mathematical models in connection with disease spreading and prevention has a long history. One of the first documented occurrences dates back to 1760, when Daniel Bernoulli used a simple model to show that the risky procedure of inoculation with smallpox was beneficial on population level. Until now, smallpox remains the only disease ever eradicated. Throughout the centuries, mathematic models have been refined and expanded, benefiting from the computational revolution. It has established itself as a field of its own, uniting a wide range of methodologies and approaches from a range of disciplines [94]. Mathematical modelling was applied to a variety of diseases including the

bubonic plague [45], measles [38, 46] and a subset of vector borne diseases [6, 107, 60]. More recent epidemics were investigated in real time and in retrospective by the epidemic models [91, 34, 22]. Classical epidemiological models subdivide a population into compartments according to their infection status, risk or age class. One of the most common epidemiological models are the SIS and SIR models, where each letter indicates a population compartment according to the infection status. In the SIS model a susceptible (S) individual can be infected (I) and returns to the susceptible compartment after recovery. In the SIR model the disease provides immunity, hence after infection an individual is assigned to the removed / recovered (R) compartment. Depending on the disease in question the model can be expanded to account for more compartments, e.g. exposed asymptomatic cases (SEIR) or partial immunity. Compartmental models can be implemented as systems of differential equations or as stochastic agent based models.

When speaking about the global spread of disease the above models have a drawback. They are designed to model the disease spread in a single population with a central assumption being the homogeneity of the population. Worldwide population is composed of several semi-isolated communities connected by air, sea and land traffic fluxes. Thus, the global population has a complex structure which needs to be accounted for if the model is to make reliable predictions. It was in fact demonstrated that models incorporating the structure of traffic connections were able to reliably reconstruct past outbreaks [25, 41]. The network theory presents a natural way to account for the structure inside and between the populations. When modelling a single structured population the individuals can be represented by the nodes and the contacts between them are captured by the links. On a global scale the nodes represent meta-populations connected by the links derived from traffic flows. The latter approach has been widely used to study past and make real time predictions about ongoing outbreaks [95, 91, 50, 34, 84, 67, 66]. Furthermore, a sophisticated modelling tool was build based upon traffic data incorporated in a global meta-population network model [18].

Nevertheless, all the methods described above rely on disease specific parameters, like the recovery time or the number of secondary cases produced by an infected. The estimates of these parameters are especially unreliable at the beginning of the epidemic due to the fluctuations and the lack of data. At the same time, deployment of counter measures at an early stage has the prospect to stop the epidemic at a low number of cases.

An additional challenge is introduced by the change in the spreading patterns of the global epidemics. Historical epidemics were spreading in wave-like patterns, starting at the outbreak origin and propagating outwards on the map. Such course was observed in past bubonic plague outbreaks, but is not observed nowadays. Many modern epidemics have demonstrated long distance jumps across big geographic distances, thus deviating from previously known behaviour. Past epidemics were well explained by diffusion models as the spreading was dominated by close distance commuting. An infected individual was likely to die or recover before covering a long distance and carrying the disease to a remote destination. In 1869, for example, a journey from London, England, to Melbourne, Australia, in 59 days was considered an extraordinary achievement [1]. Today a non-stop flight can cover the distance in 17 hours. This drastic reduction in travelling times makes it possible for an infected individual to travel to a different continent before he recovers or shows symptoms of the disease. On par with accelerating the disease spread, the fast and accessible long distance travel also removes natural boundaries, like oceans, which prior separated populations and prevented certain diseases from reaching other continents. At the same time geographically close locations with poorly developed infrastructure can require a longer travelling time than remote destinations accessible by plane.

Due to globalisation the geographical distance becomes decoupled from the travelling time, thus changing the transmission dynamic. This has important implications for the pre-

diction of epidemic spreading, the estimation of risks and the design of countermeasures. An infection can escape a geographically remote origin via air traffic. Consequently, reliance on geographical notion of distance can contribute to a perception of false security and lack of action in situations which pose a considerable risk. Modern epidemics do not exhibit wave-like patterns on a geographical scale, thus it is not the driving force behind global spreading. Recently it has been shown that wave-like spreading patterns can be retrieved when locations are remapped according to their network properties as opposed to geographical coordinates [17]. Brockmann *et al.* propose the concept of *effective distance*, which is a modified variant of a shortest path connecting two nodes. According to the effective distance the separation between two locations a and b is proportional to the probability that an individual departing from the location a will proceed to location b. This probability, in turn, is defined by the proportion of flux going through the respective link or set of links if nodes a and b are not directly connected. When nodes are remapped according to their effective distance from the origin, the epidemic front and the wave-like pattern is restored. This was verified through multiple real-world epidemics with effective distances constructed based on air traffic between airports. The study has demonstrated consistent results emphasising that long distance traffic has a similar predictive power as geographic distance had centuries ago. Brockmann *et al.* also demonstrate that the effective distance to the correct origin only results in the wave-like pattern, thus the methodology can be used to retrospectively reconstruct the outbreak origin.

Even before networks were considered in epidemic context, dynamic processes were studied on them. Analysis of random walks is one of the topics which provided multiple concepts now widely used in network based epidemiology [63]. An important property defined in this context is the *centrality* of a node. It can be defined in a variety of ways and thus emphasise different aspects of node's localisation. Centrality ranges from simple definitions like the degree centrality, which indicates the number of direct neighbours, to complicated and computationally heavy metrics. An often utilised centrality measure is the betweenness centrality, which indicates what fraction of shortest paths crosses the node in question. Consequently it indicates how many shortest paths are disrupted when the node is removed from the network. Several papers have applied these centralities in context of epidemic spreading and countermeasure design. In the former case, the spreading potential of a node was studied in connections with its centrality characteristics [51, 28], in the latter case centrality measures were used to identify targets best suited for the application of countermeasures, e.g. vaccination [39, 99].

The disadvantage of these centrality measures is their independence of context. Betweenness centrality, for example, is calculated from paths between all possible pairs of nodes, while for a specific outbreak shortest paths from the outbreak source play a more important role. Hence, conventional centrality measures assign the same importance to a node regardless of the scenario at hand. Nonetheless, we can easily imagine a scenario in which a hub is central to a specific region, but plays little role in case of an outbreak outside of this region. By considering a non-context sensitive metric the importance of a node can be overstated in some cases and underestimated in others, leading to the deployment of intervention at wrong locations and a wrong time. As will be demonstrated in this work and has been in part shown in [17] the role of a node varies greatly in different outbreak scenarios. Outbreak sources which are geographically close can lead to very distinct spreading pattern with transmission over different routes. A metric which disregards the outbreak origin is unable to capture the differences. To overcome this limitation a context sensitive metric is needed. In this thesis, we propose two centrality measures which account for the context of the outbreak and assign a node the importance characteristic for a specific outbreak. The metrics are based on effective distance trees introduced in [17], thus they rely on the information about

the origin of the outbreak.

When the source of the outbreak is known and important nodes are identified, deploying informed interventions is still a challenge. When it comes to the design of counter measures the decisions are often based on 'common knowledge', political ties and relations, or on the results of elaborate simulations. The latter are only possible several months into the epidemic, after sufficient number of cases have been accumulated and reliable estimates for disease parameters are available. Consequently, informed countermeasure deployment is only possible at an advanced state of the epidemic and thus many opportunities to control the epidemic growth are missed. At times when elaborate models are unavailable decisions are often guided by geographical distances. It has been outlined above how this can lead to false conclusions and inappropriate responses. Especially troubling are the outbreaks in geographically remote locations, e. g. multiple African countries. These often have poor economies, a fragile infrastructure and an unstable political situation. As has been demonstrated by the Ebola epidemic in 2014 underdeveloped health system, bad sanitation, insufficient surveillance and lack of education can lead to epidemics spiralling out of control very fast. In such cases an early intervention is essential to prevent the spread of the disease and save many lives, but it is often missed or considered on a much later time point due to the underestimation of the threat an outbreak poses globally.

Even when international awareness of a specific pandemic is present, the decision about the right resource allocation is not trivial. The possibilities range from allocating supplies and personnel, to the affected region to ramping up preparedness inside of yet unaffected countries. The former is usually perceived to be an altruistic, pro-social act, while latter is seen as selfish. Consequently depending on the entity owing the resources different strategies are employed: the WHO and multiple NGOs employ the 'pro-social' strategy, offering help to the affected regions. The international community is rather reluctant, providing support to the outbreak source mainly in cases of humanitarian crises like the Ebola pandemic. Nonetheless, this classification of strategies is based on common knowledge rather than research. There is a body of work on the resource allocation topic, both as a post deployment review or as purely theoretical work. The review of the effectiveness of the intervention in retrospective is difficult to interpret as no controlled conditions can be guaranteed to make the evaluation reliable. As for theoretical work, different modelling and analytical approaches were employed to study this topic. All but one study on global resource allocation aimed to minimise the global cost, thus assuming that the global good is the state desired by the allocating entity. While a variety of the studies have been done in human epidemiology [87, 86, 85], studies from the field of computer networks can also be conclusive for pandemic research [37, 117]. Resource allocation on a more local scale is well represented by human vaccination behaviour, where a vaccine is the resource distributed among individuals to prevent the spreading of the disease in the population. One of the important questions asked is how a certain population can be protected by a minimal amount of vaccines. This was extensively investigated with and without the application of game theory with regarded for the decisions of selfish individuals [32, 33, 12, 26, 83, 97]. Motivated by the global spreading the question of resource allocation between a pair of coupled meta-populations was studied in multiple papers [64, 93, 103].

While the countermeasure deployment is not a public good game in its original form, the two processes share multiple features. All nodes in the network benefit from implemented containment efforts while only the donor of the resources bears the cost. Thus, there is incentive to free ride. This has been demonstrated in context of vaccination behaviour, where according to theoretical results in a system of selfish agents herd immunity can never be reached under voluntary vaccination conditions [12]. From the public perspective herd immunity is the desired state, hence a discrepancy between the desired global and selfish

optima arise. This has been described as a general feature of public good games [73].

In most of the papers studying resource allocation in context of global epidemics the final resource distribution is imposed on the nodes by some external entity which distributes either limited resources or tries to achieve a given goal with minimal possible resources. In the real world most of the resources are owned by countries and hence no external strategy can be imposed on the usage of these resources. While after implementation of the IHR the WHO gained some directive authority, there is no legal instrument to punish the states deviating from the proposed line of actions. As has been demonstrated by game theoretical studies, cooperation in public good games is difficult to maintain in the absence of punishment [30]. Therefore, the resource allocation decisions in the real-world are best described by a system of agents who individually make allocation decisions and not a global entity imposing the distribution rules in a top-down manner. When investment decisions are left to individual agents different behaviours can be expected: pro-social, which aims for the good of all, selfish, which regards only the own cost, and a variety of intermediate predispositions. Another lesson learned from the field of game theory is that in public good games the globally optimal solution differs strongly from the optimal outcome for selfish individuals. Thus, there is reason to assume that the same holds for the investment behaviour in the case of a pandemic. Wang *et al.* has investigated the resource distribution in two connected meta-populations from a selfish and global perspective [103]. It was demonstrated that there is in fact a deviation between the global optimum and the solutions optimal to either of the infected. Consequently, in other investigated systems we expect a different equilibrium to arise when the resource allocation is optimised by individual selfish agents.

During a pandemic the decision making is heavily influenced by the perceived risk posed by the outbreak. The risk of case importation is central to the estimation of the overall threat of the current situation. It can be evaluated by elaborate epidemic models, but as outlined above precise predictions are only available after an extensive time period. We argue in favour of a more readily available method for risk estimation that enables to evaluate the threat at an earlier stage. Import risk, the driving force behind the resource allocation decisions, is highly dependent on the network topology, thus we expect the latter to have an impact on the resource allocation as well. It is important to investigate whether certain topologies can facilitate globally optimal decisions even when the optimisation is performed by selfish agents. When risk assessments are made with geographical distance in mind some locations are less likely to receive help due to their perceived remoteness. We will demonstrate that the same nodes can be well connected if the distance is calculated according to the network topology. At the same time other nodes can be at topological disadvantage, leaving them at lower probability to receive investment. Previous theoretical studies have investigated systems subject to a single threat or disease. In recent years concurrent crises were observed multiple times, hence decisions on countermeasure deployment and resource allocation are made while the international community is facing multiple potentially global issues. For that reason it is necessary to understand how the the number of outbreak sources influences the decision making and the final distribution of resources.

In this thesis we approach the topic from a network based perspective and define the necessary methods before concentrating on the resource allocation problem. We use the effective distance trees previously defined by Brockmann and Helbing in [17] and introduce a new centrality measure based on these. These metrics describe the role of a node in case of a specific outbreak showing which nodes are most likely used by infected individuals as transit location and are hence best suited for countermeasures like passenger screening. They allow to judge what fraction of the global population benefits from protective measures deployed at the node in question and how a change in the location at which the countermeasures are deployed changes the protective effect. Further we derive a method to evaluate the risk of

case importation into a specific country given a distinct outbreak origin. We demonstrate how this metric can be used if the information about the origin is missing or is imprecise. Using this risk estimation we device a model to study resource allocation on a network of countries and report its generic features like the emergence of self-investment or resource allocation to the source of the outbreak even in case of selfish optimisation.

As has been previously outlined the world aviation network (WAN) plays a central role in the spread of diseases on the global scale. We have used it to capture the interactions between different locations, i.e airports and countries, during pandemic situations. In all the work described on subsequent pages, WAN and transportation networks as a general class play a central role. On the WAN nodes represent the airports which are connected by air traffic and are measured in means of passengers on commercial flights. These are the carriers of the disease in case of a pandemic and long distance flights enable them to travel to remote locations during their infectious period. It has been demonstrated that the reliability of global models increases when the air traffic network is accounted for and therefore we believe that it has to be at the core of a metric aimed at the prediction of global pandemic spread.

As the first step, we define two centrality measures: scope and confluence. Both metrics are calculated based on the effective distance tree derived from the original network. Hence, both metrics evaluate the role of the nodes based on their presence in the most probable paths from the outbreak origin to remaining nodes of the network. *Scope* of a node n represents the fraction of the network downstream of n on the effective distance tree, thus the population which can be reached through a shortest path traversing the node in question. The initial cases imported into the country downstream from the node n are likely to be brought in by traversing node n and hence the downstream nodes benefit from the countermeasures implemented at n, such as passenger screening. We call a node starring a high scope value a *gate*. The second metric proposed in this work, *confluence*, indicates the branching of the tree after the node in question. Consequently, a node with high confluence represents a bottleneck on the effective distance tree which is a part of a high fraction of shortest paths and the latter are diverging after passing this node. Such bottlenecks are valuable targets for countermeasure deployment as they present a good (and last) opportunity to efficiently protect a high fraction of population. Any shift of containment measures to a further downstream node will lead to a reduced protective effect. In cases when the outbreak origin is uncertain or is stretched over a great geographic distance each metric can be averaged to account for the situation at hand. We will explain the procedure and outline the benefits and downsides of it. Both metrics were motivated by the application in a pandemic scenario, hence we demonstrate the distribution of these metrics across the WAN. Using scope and confluence we characterise the nodes of the WAN in terms of their role in a variety of outbreak scenarios. We establish that a role that one node can take on varies strongly depending on the outbreak situation: a node can change from being a major gate in one scenario to be a leaf node in another case. Taking a global average of the metric conceals this highly variable role, thus leading to similarly wrong conclusions as established centrality metrics. To demonstrate how deliberate averaging can be used to aid decision making, we apply scope and confluence on a selection of real-world and hypothetical outbreak scenarios spanning extended geographic regions. We identify important global and regional gates in each case. Using scope averaged over 22 distinct geographic regions we define a profile of an airport and show that it bears important information about a particular hub. The profiles highlight functional differences between the airports in terms of their region of influence and describe how often a node acts as a gate.

Relying on the effective distance trees we devise a method to estimate the probability of case importation to a specific node during an ongoing epidemic. It regards an important feature of the WAN, i. e. the existence of the hubs used for transit flights and the smaller

airports which only serve as terminal targets. The feature leads to a profound difference in probability that an agent will exit the transportation network at a specific node. As the risk of case importation is highly dependent on the exit probability and the information about it is rarely available we establish a method to infer the exit probability from the topology of the effective distance trees. Just as in the case of scope and confluence, import risk is not dependent on specific disease parameters. The only information required is the origin of the outbreak or the currently affected region, thus making import risk a measure available on early stages of the outbreak. To demonstrate the application of import risk we use it in a set of real-world and hypothetical scenarios. Similar to the metrics described in the paragraph above, import risk can be averaged over a broad geographical region to account for the spread or other factors of the disease. The weighting during the averaging procedure can reflect the specifics of the location or the severity of the outbreak in the target location. All in all we show how import risk can be used to evaluate the threat of an outbreak and predict locations which are likely to experience early imported cases.

Using the import risk we develop a simple model to study the resource allocation problem. Considering the benefits of a network based approach the defined model incorporates the actors as nodes of a network. The actors represent countries connected by traffic. We develop the model with a network analogous to WAN, we do not limit the exact kind of traffic. To reduce the complexity and enable the understanding of basic laws this model is based on an artificial, synthetic network. As the network is the core part of the model and plays an important role for the distribution of import risk across the nodes we investigate how changing topology affects the decision making by the acting agents. The model introduced on the subsequent pages aims to understand the resource distribution optimal from the global perspective and the perspective of individual selfish agents. The proposed model incorporates that the resources are held by the individual countries. Hence, each node is an entity owning resources and making decisions about how to best distribute those in the face of a current pandemic. We call the final distribution of resources of a single node its strategy. In the thesis at hand we study two different scenarios: a fully pro-social network, which is equivalent to optimisation by a global benevolent entity, and a network of selfish players, who minimise solely their own cost. In both cases the optimisation is an iterative process consisting of multiple rounds of decision making and revision. As expected from previous knowledge there is a discrepancy between the outcome of both optimisation modes. We demonstrate that the results in fact deviate, but only to a small extent for many cases. In outbreaks originating from a single location or a set of adjacent locations the globally and selfishly optimal solutions differ only in the amount of donated resources, but not in their distribution. This situation resembles the conflict demonstrated by Bauch *et al.* in case of voluntary vaccination [12]. Our model demonstrates that resource allocation to the source of the infection is the optimal solution in both cases. We present analytical evidence pointing in the same direction. To investigate how the decision making is changed by increasing number of outbreak sources we implement the multi-source outbreak scenario. Here the discrepancy between the optimisation modes increases, giving rise to a strategy of self-investment in the selfish scenario. Nodes in close proximity of the source of the outbreak profit most from investment in the source even in selfish optimisation scenario. The nodes threatened by multiple sources to the same extent adopt self-investment as the dominant strategy. Our model demonstrates that the driving force behind the investment strategy is the probability of case importation from the infection sources present in the network. In case of multiple outbreak sources the ratio between these probabilities becomes of a particular interest for selfish agents.

All outlined results are stable across the tested network topologies implying that they are generic to the problem at hand. The results are also stable with respect to multiple

additional parameters of the model, which are described in greater detail in the respective chapter. While the distribution of the resources across the network does not play a role for the final resource distribution, the overall amount of resources is of high importance. Our model suggests that a selfish system with low resources is more prone to self-investment. Consequently, it might drive the system into a vicious cycle, where affected countries are left unaided during the epidemics, their economies are damaged, thus resulting in an economically weaker system. Our research also highlights how benevolent organisations can influence the final resource distribution of the system by making informed investment into the affected states.

This thesis is structured as follows:

- Chapter 2 gives an outline of approaches available in the field of the epidemic modelling, highlighting the methods borrowed from the network science and game theory; it gives a broad overview about the current theoretical methodology for risk estimation and the most recent research about countermeasure deployment and disease containment.
- Chapter 3 summarises the research which serves as a basis for the work presented in subsequent chapters; in gives a detailed description of the world aviation network, its features and further summarises the concept of the effective distance introduced by Brockmann and Helbing in [17].
- Chapter 4 introduces the centrality measures of scope and confluence; derivation of the metrics, their comparison with established metrics and application is presented in this chapter.
- Chapter 5 defines the import risk and exit probability, presents a detailed mathematical definition and demonstrates the application of it in hypothetical and real-world scenarios.
- Chapter 6 describes the model derived to study the resource allocation problem; here we present the mathematical definition and thorough analysis of the model on different topologies and in a subset of special cases; we further demonstrate computational results on a variety of networks with different conditions and parameters.
- Chapter 7 summarises the results putting them in a bigger picture; we outline how our results can be used during the epidemics to come and how our methodology can be extended.

Chapter 2

State of the Art

This chapter summarises scientific findings, which lay the foundation for the thesis. As the work presented in subsequent chapters is from a multidisciplinary field, this chapter present the state of the art of network science, epidemic modelling, optimisation and game theory. Nonetheless, the emphasis lies on the application of the methods in context of human disease spreading. First, the mathematical concept of a network is introduced, followed by approaches to find shortest and most probable routes between the nodes of the network. Epidemiological models are outlined in the next sub-chapter, emphasising the differences between the findings based on classical and network based methods. Game theoretical approaches and their application to disease containment are explained thereafter. Finally, existing attempts to answer the questions approached in the thesis are summarised.

Presented work is rooted in different areas of science and requires understanding of the concepts from multiple fields. In this chapter we will explain the topics required to fully understand the work presented in subsequent chapters. While it gives a brief overview, it is not a full review of the topics. We refer an interested reader to consult the literature referenced in the respective chapters. Finally, we will introduce multiple papers, which which are concerned with a similar topic as explored in this thesis. The differences in methodology applied and the questions posed will be discussed in detail.

There are two building blocks of the model. First is network theory. Chapters 4 - 6 require basic knowledge of network theory, which will be presented in this chapter. We will explain how graphs are constructed, which benefits arise from structuring the population of a model as a network and what methodology exist to work with graphs. Apart from the concept of a network as a structure to capture the relations, an important role is assigned to the path finding algorithm. Finding path from node a to node b is a necessary requirement for answering multiple questions. This is especially true for spreading phenomena, which use the links as their spreading routes. As will be explained below, finding a path is a non trivial and often computationally intensive problem. We will explain the most established method to find shortest paths on weighted networks, the Dijkstra algorithm. Based on it, the Yen algorithm enables the user to determine k shortest paths instead of only one. We will outline modifications which make the latter possible. We will briefly explain how the Yen algorithm can be modified to find all paths with length above a certain threshold. Finally, we will give a brief overview of successful application of networks in the disease modelling and point out the milestone findings.

Second building block are concepts borrowed from the game theory. The cost function optimisation plays central role in game theoretical research, as it does in our model presented

in chapter 6. Hence, we will illustrate the basic approach to function optimisation in general. Further we will describe the evolutionary strategy method of optimisation, which is at heart of the implementation of the model proposed in chapter 6. While we did not use the classical version of the evolutionary strategy algorithm, understanding of this approach is required to fully understand the modified version used in this thesis. In the last section we will review research centred around the resource allocation topic. Multiple papers were published in recent years with the aim of solving this important question. While concerned with similar question as our research, the methodology and the basic assumptions differ. Most central difference is the entity distributing the resources and the cost function being optimised. All but one paper discussed in this chapter assume an overarching actor, who distributes the resources across the network to achieve an outcome with the minimal cost for the entire network. We will discuss the implications of these assumptions in more detail in the last section of the chapter.

2.1 Basic network theory

Recently, graph and network theory experienced a rise in popularity and were incorporated into studies across different fields. Among these are the social sciences and, more recently, public health and epidemiology. Both fields benefit from the introduced methodology and contribute to the development of graph theory itself. Hence, there is a mutual interest for both research areas in a joined effort. The work at hand follows along this line, using network theory to study a public health related question. Networks present a very convenient tool to describe a structured population or a group of agents sharing distinct relations. They allow individuals in a population to be distinct in some features, while sharing others.

A most simplistic network is composed of a set of nodes, with the number of nodes usually denoted by N. These nodes are connected by links, where the number of links is denoted by L. Both, the nodes and the links, can have properties assigned to them. A most common property of a link is its weight, which represents the strength of a link. Usual notation of a weighted link from node n to node k is w_{kn}, where n is source and k is target. There is no consensus on how a link weight is defined as the definition heavily depends on the objective for which the network is constructed. Multiple kinds of link weights exist. On the one hand, a weight of a link can represent the distance or the cost, e.g. the distance in terms of kilometres in a road network or transport fees of a rail network. In such a system two nodes are farther apart if they are connected by links of greater weights. Such definition is appropriate if travelling times or resource spending are the matter of interest. Another way to define a link is to associate its weight with a capacity or a flux possible via this connection, e.g. in case of passenger numbers on a specific traffic route or capacity of connecting pipes. When this definition is used, a location is closest when it is connected by a link with a high weight. Regardless of the definition, the distances over multiple links are calculated by adding the weights along a specific path. The sum of the link weights is the length of the path. The shortest path according to the distance-based notation is the one resulting in the smallest possible sum of links. Using the flux inspired definition, the path responsible for the biggest flux between target and destination is the one resulting in the biggest sum along its links. Note that the usage of the word length in the latter case is rather misleading. The exact procedure of finding the shortest path is described in the subsequent chapter. Apart from the weights, links can be assigned a direction, thus $w_{nk} \neq w_{kn}$. Directed links can arise on road network when one way streets are present or in a biological regulatory network, when one gene is inhibited by product of the other, but not vice versa. Some of the methodology described in the subsequent sections can not be applied on weighted or directed networks.

We will explicitly state the limitations of each method.

A property can be also assigned to nodes to characterise them in a greater detail. Properties can be descriptive, e.g. characterising the kind of a node or its affiliation to a certain group. Other node properties can arise from its position on or features of the network. Most basic property of a node n is its degree, k_n. A degree describes the number of links attached to a node. In case of an unweighted, undirected network, a degree also represents the number of neighbours of the node. In weighted networks a weighted degree can be defined, $k_n = \sum_j w_{jn}$. The latter accounts for the strength of the links, thus using more of the available information. On directed networks, two different degrees can be defined: in-degree, which accounts for the incoming links, and out-degree, which describes the outgoing links. Especially in social sciences degree is used as a centrality measure, it shows the influence of a node when links represent social contacts or friendships.

There are three additional centrality metrics, which will be used in this thesis. First is the closeness centrality, it describes the maximal distance of a node to any other node in the network, $C_c(n) = N/\sum_k d_{kn}; 1 \geq C_c \geq 0$ where d_{kn} is the length of the shortest path from n to k. Defined in this fashion, the closeness centrality approximately describes the average path length from the node n to every node in the network. While in undirected networks closeness centrality is equal regardless whether the paths to or from n are used. In the directed networks, a node can have high closeness centrality for the incoming, but low for the outgoing paths and vice versa. As per convention two nodes n and k which are not connected by a path have $d_{kn} = d_{kn} = \infty$. Thus, for such nodes $C_c(n) = 0$. To prevent this, the sum of the quotient is used instead of the quotient of sums, i.e. $C_c(n) = \sum_k 1/d_{kn}$. The latter definition was proposed in multiple publications under the name of valued or harmonic centrality [92, 29].

The second centrality measure is betweenness centrality, which reflects how often a node is part of a shortest path. Betweenness centrality is defined as $C_b(n) = \sum_{ts} \chi^i_{ts}$ where $\chi^i_{ts} = 1$ if the shortest path from s to t traverses i and 0 otherwise. Hence, betweenness centrality counts the number of the shortest paths traversing node n. A node with a high betweenness centrality is crucial in connecting the nodes of the network. A disruption of this node, e.g. its removal, will have a big impact on the traffic in the network. To make betweenness centrality more comparable across networks of different sizes, it can be normalised by the number of all possible paths

$$C_b(n) = \frac{\sum_{ts} \chi^n_{ts}}{N(N-1)}$$

Latter definition is used in chapter 4. Betweenness centrality has been widely used to estimate the role of a node during pandemic spreading, it was suggested that nodes with high betweenness centrality are valuable targets for vaccination [28]. Similar arguments were made in connection with the degree of a node. Immunisation with regard to centrality metrics has shown better results than the pure random immunisation. Nonetheless, high amount of nodes have to be immunised to achieve a transmission disruption. For more details about the application of centrality measures for pandemic prevention and disruption we refer interested reader to one of the following publications [28, 39, 42, 51, 62].

In recent years advances in technology enabled us to record a variety of the real world networks: transportation, social contacts and regulatory networks. Multiple synthetic network types were created to resemble their natural counterparts. There are some features, which are often observed in nature and are generic for many real world networks. We will outline the evolution of synthetic network types and explain how introduced features change general properties of the network.

One of the basic types of a network is a chain. The chain is a collection of N nodes where

each node with index n is connected to nodes with index $n-1$ and $n+1$, if such a node exist. Following from the topology, there is only one path from each node to any other if each node is traversed only once. Deletion of any link will result in a disconnected network, with some parts being inaccessible for the rest of the network. A chain can be modified to represent a ring, when nodes with indices 1 and N are connected. A network with slightly more complexity is a lattice or grid graph. Such graph forms a regular tiling when represented in Euclidean space [108]. Additionally, a grid with looped boundaries can be defined, such that the nodes at the edge of the lattice are connected to edge nodes at the opposite side. Given that, all nodes on the lattice have exactly the same number of neighbours. In a lattice each pair of nodes is connected by multiple paths, even if each node is traversed only once. Consequently, this network is very robust against link deletion. Due to the regular structure of this graph, multiple shortest paths between the nodes exist. On a chain, a ring and a lattice the maximal shortest distance between two nodes, called the diameter of the network, tends to grow linearly with the size of the network. In many real world networks diameter tends to be small and increase slower with the growing network size.

An important feature of networks found in nature and technology is the small-world property. In such networks, diameter tend to be short and increases logarithmically with the growing network size [104]. The popular statement about "six degrees of separation", referring to the number of acquaintances needed to connect every two person in the world, points out the same phenomenon [70]. A network type which exhibits small-world properties is a random graph, with most prominent example being Erdős–Rényi model. For every pair of nodes n and k a link is established with probability p. Average degree of an Erdős–Rényi network is $\langle k \rangle \approx N * p$, the diameter of the network is considerably lower compared to lattice with similar degree. Note that, in contrary to lattice, the degree of nodes in a random graph follows a Poisson distribution. Furthermore, the final network can contain disconnected nodes. Multiple studies indicate that small-world networks are more stable against perturbations [105, 4, 3, 53, 52].

Another aspect of many real world networks is the presence of high degree nodes, so called hubs. These nodes accumulate many connections and emerge as the transit locations when agents or information is travelling across the network. Prominent networks starring high number of hubs are transportation networks (e.g world aviation network) and man-made communication networks (e.g. internet network). In such networks the degree distribution follows a power law with exponents $2 < \varepsilon < 3$. The network category which features the small-world property and a power law degree distribution are scale-free networks. Barabási–Albert model is one of the best known and studied models to generate a scale-free network [11]. In this model is initiated with multiple seed nodes, the subsequent nodes are added one by one. With each node l links are added. These links lead from the node in question to an already existing node. Probability to connect to a node is proportional to nodes degree, thus nodes with high amount of links are more likely to gain additional connections. In an Barabási–Albert network the minimal degree is l and the network is fully connected. It can be shown analytically that the exponent of the degree distribution is $\varepsilon = 3$. Networks of this type exhibit multiple interesting features, e.g. very high robustness against random node removal [23, 19, 4]. As the probability of randomly removing a hub node is low, networks integrity is conserved even at high attack rates. At the same time, the scale-free networks are very susceptible to targeted attacks, which eliminate the hub nodes. Furthermore, diameter of such networks increases at an even slower rate due to preferential attachment to the hub nodes.

While there is a variety of more complex synthetic networks, the types listed above already exhibit many features observed in natural and technological networks. An important example, world aviation network, exhibits both mentioned properties: the small-world be-

haviour and the heavy tail degree distribution. Moreover, this network shows a community structure, which is not part of any network types described above. World aviation network is discussed in more detail in section 3.1

2.2 Path finding, Dijkstra and Yen algorithms

As has been mentioned in the previous section, the task of finding paths from a source node to a target is not a trivial task. In general, amount of paths between two nodes is infinite, if some nodes can be traversed multiple times. Nonetheless, a path with loops can never be the shortest, hence such paths are of no concern to an algorithm searching for the shortest possible connection. The most simple and common algorithm, which finds paths from a source node to all possible targets in the network is the breadth-first search. It can only be performed on an unweighted network as it minimises the number of traversed edges. When the network is weighted, a path via many short links can be in fact shorter than a path via a few long links. Note that we use the terms applicable to distance-based links for simplicity. The breadth-first algorithm is unsuitable for the transportation networks, where link weights are available.

The methodology, which can be applied to find shortest paths on weighted networks is the Dijkstra algorithm. Similar to breadth-first algorithm, it finds paths from the source node to all the nodes in the network. Note that no target node can be specified to the basic algorithm, thus the fastest way to find a path from the source to the node of interest is to run the Dijkstra algorithm until the target node is reached and thereafter terminate the algorithm. This method is guaranteed to find the shortest possible path [70].

The algorithm is initiated with two arrays. The first array $\mathbf{d}$ holds the best current estimate of the shortest distance to the source. Note that the distance estimate has to be the upper bound for the possible shortest path length. Distance from source s to itself is set to $\mathbf{d}_s = 0$, for all other nodes $\mathbf{d}_k = \infty$ is the most farthest possible distance. The second array indicates whether the actual shortest distance was already found. At initiation all values in the second array are set to 0 to represent that distances set at initiation now are uncertain. Further the algorithm follows an iterative procedure:

1. Among the nodes, whose estimated distance is not certain, a node n with the smallest estimated distance is chosen

2. Its distance is marked as certain

3. Distances to all neighbours of n are calculated by adding the weight of connecting link to the distance of n; if the calculated distance is shorter than the current estimate for the neighbour, the estimate is replaced by the new shortest distance

4. the procedure is repeated from step 1 until no nodes with an uncertain distance are left

If the calculation is performed on directed graphs only outgoing links are considered in this calculation. Crucial step to this procedure is step two, where current estimate of the distance is marked as certain. We can be sure that established distance $\mathbf{d}_n$ at this point is the shortest possible due to following consideration. As per definition there can be no negative weight to a link. Hence in a path shorter than the current estimated distance to n, distance to all intermediate nodes have to be shorter than $\mathbf{d}_n$. No unexplored nodes fulfilling this condition remain since we have chosen n as the node with smallest uncertain distance. Neighbours of all nodes, which distance has been marked as certain, has already been explored and

compared to n's shortest distance estimate. Thus n's current shortest distance is in fact the shortest possible distance and we can safely mark it as certain.

Particular problems require knowledge of a set of K shortest paths. An algorithm, called the Yen's algorithm, was designed to solve this problem on weighted and unweighted networks [115]. To initiate the Yen's algorithm, the shortest path from the source node to the target node has to be found. To do this any of the path finding algorithm described above can be used. Assume we found the shortest path $\{1, ..., i, ...t\}$ where 1 is the source node and t is the target node. First path is stored as A^1, where A is the list holding all acquired shortest paths. If there are multiple shortest paths available, the choice of A^1 is made randomly and the remaining paths are assigned to the list B. List B holds all candidate paths. If at first step only one shortest path is acquired, list B is created empty.

For any subsequent path a different procedure is followed. To find the next shortest path k we assume that paths $1, .., k-1$ have been found.

1. To find next shortest path A^k following procedure is repeated for each node $i \in \{1, ..., t-1\}$ in the pathA^{k-1}
   (a) Assign a subpath of A^{k-1} from node 1 to i to be the root path R_i^k
   (b) For each path A^j, with $j = 1, 2, ..., k-1$, check if the the subpath from 1 to i-th node is equal to R_i^k; if so, the link between the nodes i and $i+1$ of the respective path is temporally removed
   (c) Find the shortest path from node i to the target node t; this path is called the spur path S_i^k
   (d) Root and spur paths are joined to form path A^k, leading from source to the target node; this path is added to list B
   (e) All removed links are restored and the process is repeated for the next node i, which is part of A^{k-1}
2. From set B choose the shortest path and add it to A as A^k; repeat from step 1 until all K paths are found

Note that the list B already contains a set of shortest paths, hence as soon as B contains K or more paths the algorithm is finished. We can fill the list A with shortest paths from B. As proposed by [21], a path is a shortest possible path when it is composed of subpaths, which are in themselves minimal. For detailed derivation and the proof see [21, 115].

All algorithms described previously are designed for the link weights, which are additive. We can imagine a network where the link weights represent transition probabilities. Consequently, when calculating a path length in this probabilistic notion, the link weights have to be multiplied. Both of the presented algorithms can be applied in such a network after link weights have been transformed to $w_{kn}^* = -log(w_{kn})$.

2.3 Epidemiology and disease spreading on networks

There is no consensus about what was the first epidemiological study ever performed. But multiple landmark papers can be identified throughout history. The one which marks the first use of mathematical modelling in the history of epidemics is a paper written by Daniel Bernoulli in 1760 about the smallpox and inoculation [14]. Bernoulli was arguing in favour of the inoculation technique against small pox using the mathematical reasoning and the population data available at that time. While attempts to use population wide information

were done prior to him, e.g. in the prominent cholera outbreak investigation by J. Snow in 1854 [100], Bernoulli's article was one of the first applications of exact mathematics to the matter.

In context of classical epidemiology, no population structure is modelled. A common assumption is that the population is well mixed, thus each individual has the same probability of a contact to any of his peers. While not reflecting the details of reality, it is a valid assumption when the data about the population structure is unavailable. Consequently, an assumption about the homogeneity of individuals in the population is made. The only distinction is based on the infection related status. Accordingly, the population can be split into subclasses of identical individuals. Given this approach, two parameters are needed to characterise the disease: the infection rate β and the recovery rate γ. The latter is defined as the inverse of infection period. The simplest model contains only two classes: susceptible and infected, hence it is called the SI model. It can be described using only one parameter, β, as in this model infected never recover. While this reflects the properties of some infections, it is highly unrealistic for others. SIS and SIR models resolve this limitation by introducing a recovery process. After the recovery, the individuals either return in the susceptible group (SIS model) or form the recovered / removed compartment (SIR model). Multiple more complex models have been formalised, e.g. by introducing an exposed compartment, which accounts for infected, but not yet symptomatic individuals. Furthermore, the population can be divided into age groups to account for different disease dynamic among age groups. All of these models can be implemented in terms of a system of ordinary differential equation or as a stochastic system. For more details and the differences in the outcomes of these approaches refer to [47].

Using this models fundamental insights into pandemic behaviours were gained. One of the central parameters of theoretical and applied epidemiology, R_0, is derived from the parameters described in the above paragraph. This value reflects the number of secondary cases produced by a single infected in a fully susceptible population and thus can be established approximately during an outbreak. One of the most fundamental laws, highlighted by the epidemic modelling and reasoning alike, is the existence of an epidemic threshold at $R_0 = 1$, below which no disease spreading can be sustained. At this threshold an infected, on average, produces 1 secondary case. Thus, the disease can be remain persistent in the population in a deterministic system. At lower reproduction rate the number of infected declines and the disease goes extinct. Furthermore, the existence of herd immunity and the exact fraction of vaccination fraction can be derived on the basis of epidemic models. A comprehensive review of the matter is presented in [31].

Findings described above rely on the assumptions made. One of the heaviest assumptions is the well-mixed population structure. With recent advancements in technology, many ways of inferring the population structure have been created. It can be pooled from the online social networks, the mobile data and from portable sensors. The network science is a valuable tool to incorporate this information into epidemic models. It has been shown that the population structure can have a strong impact on the final size of the epidemic and the speed of its progression. As demonstrated by [47], network topology can alter the final size of an SIR epidemic. Erdős–Rényi networks exhibit the highest peak incidence and it is also reached fastest. The arising curve is comparable in its form to the classical well-mixed simulation. Similar curves are derived from other networks, which exhibit distinct small-world properties and small diameter. Other examples of such networks are geometric and scale-free networks. Regular networks, like lattices, and certain implementations of small-world networks exhibit a much slower epidemic progression and local wave-like pattern of the spreading. When the regular lattice structure is loosened by random rewiring the spreading progresses faster. This effect is due to the possibility to cross into an unaffected part of the network. The study

of epidemic spreading on networks has challenged the existence of the epidemic threshold described above. As has been demonstrated by [4], releasing the assumption of a homogeneous population alters the values of the epidemic threshold. Albert *et. al.* demonstrated that in a theoretical scenario of an infinitely big scale-free network, an epidemic threshold is non existent. Hence, if the underlying population structure permits it a disease with lower R_0 can spread.

To model the disease spreading on a more global level a combination of the network and ODE based approach can be used. In this case the network represents a set of connected meta-populations. Each meta-population is then modelled by a system of differential equations, the links indicate a coupling between the populations. This model allows to account for known global structures, e.g. air traffic connection between the countries, and include spreading of disease in a local population with an unknown contact pattern. Such model was used in multiple publications, e.g. [25, 86]. One of the most sophisticated tools to model worldwide disease spread, GLEaMviz, uses a combination of spatially separated meta-populations coupled by traffic connections. It was applied in connection with multiple outbreaks, most recently to forecast the spreading of Ebola during the 2014 outbreak [84, 34].

A different approach is taken by [17], where the network structure of air traffic is used to predict the spread of the disease without regard to any disease specific parameters. It has been shown that arrival times can be reliably predicted and outbreak origins estimated using this methodology, called the effective distance. The effective distance is described in more detail in chapter 3.

Alongside with the disease spreading on networks, the countermeasure deployment was widely tested. In network science multiple metrics describing the role of a node are known. Their usefulness in context of epidemic containment was investigated by several researchers. The main insights are summarised by [79]. Many papers highlighted the usefulness of the degree and betweenness centrality for the identification of influential spreaders [4, 79, 81]. It has been shown that vaccination strategies based on knowledge of the properties of a node are far more effective than random immunisation of individuals, requiring a lower fraction of the population to be vaccinated for the disruption of disease spreading. Other publications emphasised that depending on the spreading process classical metrics, like degree or betweenness centrality, are less suited to evaluate the role of the nodes [51, 28, 62]. During a pandemic the structure of the affected network is often unknown, as are degrees and centralities of individuals composing a network. Thus, methods to infer these from the local information were proposed and applied in context of the epidemic containment [24, 39]. Network specific interventions, like link deletion, were tested in context of disease containment [84].

In general, network science has proven to be a valuable tool to investigate a spreading processes, led to new insights and approaches. In recent years many real world networks were used to forecast the course of ongoing and emerging epidemics and to reconstruct past outbreaks [34, 66, 91, 102]. It became a common practice to evaluate the threat of an outbreak by extensive simulations, often involving transportation networks to infer the spreading routes. An extensive computational tool, using transportation networks on multiple level and meta-populations, was developed to enable interested parties to use computer simulations without programming knowledge [18].

2.4 Game theory and disease containment

Game theory is a diverse scientific field, which was applied to many different questions in economics and social sciences. It gave rise to the concept of *homo economicus*, now well

known and widely accepted in finance. Game theory was later introduced into biology by John Maynard Smith and eventually became one of valuable tools for theoretical study of evolutionary biology. More recently game theoretical methods found their way into other fields, e.g. psychology, computer science and public health.

At the basis of the game theory are players who compete in a game, trying to maximise their payoff. A game is defined by a set of rules, which in turn defines the available actions and the possible outcomes. The payoff of a strategy of a player depends on the action of its peers. This is a unique feature of game theory, setting it apart from the field of classic optimisation. Hence, actions of others have to be considered during the optimisation in game theoretical context. The players are considered to be strictly selfish, minimising their own costs. In evolutionary game theory this is easily explained by the selection pressure applied on individuals and groups. Thus, a species that suffers lowest challenges will be favoured by natural selection. In this context, even altruism can be explained in terms of selfish or group-centric thinking. We refer interested reader to [72] for further details on the game theory in evolutionary context.

There are multiple standard games, which have been investigated and solved analytically, e.g. zero-sum games or the prisoners dilemma. These games can be rephrased to fit multiple scenarios and used to investigate a big variety of questions. A game of particular interest for public health and pandemic phenomena is the public goods game described as following. At the beginning of the game all players have the possibility to contribute to a pool of public goods. After that, donated amount is multiplied by a factor and split equally among the players, regardless of their contribution [54]. In this context, free-riding emerges as a dominant strategy: a player can generate payoff without inflicting any cost to himself. It has been shown that no cooperation emerges in the classical public goods game if the players are rational [5]. Nonetheless experimental results suggest a different behaviour when this topic is investigated experimentally [54, 7]. Multiple mechanisms were proposed to explain how cooperation can emerge and be sustained, e.g. through altruism, reciprocity or punishment [30, 48, 5, 57]. Some papers suggest that parallel public good games in structured populations, e.g. networks, lead to higher cooperation than in a well-mixed group [96]. For more information about game theoretical formalism on networks see [43]. Note that in this classical definition all player have the same circumstances, thus same strategy results in the same payoff for all players. The assumption that all players are equal, is not fulfilled in a variety of situations, e.g. in the model proposed in chapter 6. However, a broad spectrum of questions can be addressed by formulating those as classic public good games.

The vaccination behaviour, vaccine acceptance and levels, achieved by voluntary vaccination are highly relevant topics for public health and policy making. In recent years it was widely explored using the game theory. The phenomenon of herd immunity can be well framed in the public good terminology. For any disease eradication achieving herd immunity is necessary, thus it is crucial to understand the mechanisms which promote vaccination. The vaccination can bear cost for an individual in terms of time, money and adverse effects. On the other hand it prevents cost inflicted by infection and disease. With an increasing fraction of population being vaccinated, the incentive for an individual to vaccinate declines as infection becomes less likely. Using game theoretical formalism, [12] showed that herd immunity can never be reached under voluntary vaccination policy with selfish, rational agents. This also requires a player to have complete information and to explicitly evaluate the payoff of a strategy. More realistic optimisation strategies with limited information and learning behaviours were investigated, but led to globally suboptimal outcomes like temporal or spacial flare-ups of disease [26, 65]. When, however, a network structure is introduced into the game, a different outcome arises. As shown by [83], when disease is spreading over a contact network and the vaccination behaviour is dependent on local information eradication of

disease can be achieved even when the policy is voluntary.

As has been mentioned in the paragraph above, different algorithms of cost optimisation can be used in context of game theory. In the subsequent section we will explain challenges and basic methodology of cost function optimisation. Further we describe the basic algorithm used in chapter 6 in detail.

2.5 Optimisation and evolutionary strategy

As explained in the previous section, the cost function optimisation is a central part of game theory. In simple two player games solutions can be found by reasoning or by back of the envelope calculations. With an increasing number of players and complex interaction networks finding the final equilibrium is not simple. For more complex systems a complete, analytical solution can often not be found. Thus, an iterative heuristic must be used to acquire the final equilibrium of the system. A variety of algorithms can be used for this purpose. Depending on the function, which is being optimised, multiple approaches exist. A very popular method is gradient descent. In this method, the gradient of a function at the current position is evaluated and this information is used for the most efficient descend. While intuitive, this method is problematic when a function has multiple optima, is not differentiable or must be optimised over a large number of parameters. Discrete parameter values can be a further source of problems.

To solve problems of this kind, a family of algorithms inspired by the evolution arose. Three distinct methods exist: evolutionary programming, genetic algorithms and evolutionary strategies. All these algorithms incorporate two opposing forces, the diversification through randomness and the pursuit of optimality through selection. In each case the optimisation is an iterative process, where at every step a set of possible parameter configurations of the function is created and subjected to selection. Each parameter configuration in the set has a measure of quality, called fitness, upon which parameter configurations are selected. The selection process is described in detail at the bottom of the paragraph. The set, which remains after the selection process, is called a generation. From each generation new individuals are created by a process which mimics the mating. During this process a new parameter configuration is created as a copy or a permutation of multiple parameter configurations from the previous generation. The template configurations are called parents and the created ones are called offsprings. Additionally, the mutation of offsprings can be performed to introduce a greater variety and explore the parameter space. After this step parameter configurations with the highest fitness are selected to form a new generation and the rest of the set is discarded. In this thesis we use evolutionary strategy for the cost optimisation. We will explain this methodology in greater detail in subsequent paragraphs. We follow the procedure described by [16].

Before describing the algorithm some basic concepts have to be introduced. In this algorithm an individual is described by a parameter configuration for the optimised function and the endogenous parameters of the individual. The latter affect the mutation rate of all properties of the individual as the evolutionary strategy algorithm allows mutation of function and endogenous parameters. Assume, for example, the function we are minimising is $y = x_1^2 + x_2^2$. Then an individual is described by (1) an array of length two, $\mathbf{x} = (x_1, x_2)$, (2) a vector of endogenous parameters $\mathbf{s}$ and (3) the fitness associated with the parameter configuration, $F(\mathbf{x})$. More generally, an individual i_k is described by $i_k = (\mathbf{x}_k, \mathbf{s}_k, F(\mathbf{x}_k))$ where $\mathbf{x}_k$ is a vector of length n containing all parameters of the optimised function. The evolutionary strategy algorithm has application beyond simple numerical functions, for more detail we refer to [16]. The fitness $F(\mathbf{x}_k)$ is the measure of how close the value of the optimised function,

parametrised by $\mathbf{x}_k$, comes to the desired outcome. Depending on the implementation,

the fitness can approach zero when reaching optimal configuration or it can be maximal at the desired best configuration. In previous example we can define the fitness as $F(\mathbf{x}) = y(\mathbf{x})$, thus fitness will decrease as it approaches the optimal solution. In this section we call parameter configuration closer to the optimum to be fitter, regardless of the underlying implementation of the fitness function. The exogenous parameters, meaning unvarying parameters of the algorithm are μ - size of the parent population, λ - size of the offspring population and ρ - mixing number, determining how many parents an offspring has. The general procedure of the algorithm is as follows

1. Initialise the system by creating the first generation
2. For each offspring $k_n; n \in \{1, 2, ..., \lambda\}$
   (a) Pick ρ individuals to be parents of k_n
   (b) Create the offspring by recombination of parents' **s** and **x**
   (c) Apply mutation on the offspring; note that mutation parameters **s** have to be mutated first, followed by mutation of **x**
3. Select the fittest individuals to form the next generation
4. Repeat steps 2 to 4 until the final generation is established; this can be triggered either when a specific number of generations is reached of if the composition of the population is not changing over multiple generations

At initiation of the algorithm the first generation is created randomly. Depending on the constraints and the system, the range of random number generator can be limited. The first generation contains μ individuals, which will serve as parents. Note that the number of parent individuals is not limited to two. In case of $\rho = 1$ no recombination is possible in the next step, thus usual choice is $\rho \geq 2$. Parents are chosen with uniform probability with no regard to fitness, consequently selection is not applied at this stage. This is an important difference between genetic algorithm and evolutionary strategy, where in the former selection is applied during the mating process. In evolutionary strategy non-optimal parameter configurations can still serve as templates to create offsprings. Thus space around suboptimal solution is explored and diversity is not lost at early stages. After all ρ parents were selected, one offspring is produced as a recombinant of those. Two types of recombination can be applied to generate an offspring. The intermediate recombination produces an offspring for which each parameter is the average value of its parents. Thus for an offspring k and set of its parents $i_n; n \in \{1, 2, ..., \rho\}$

$$k_m = \frac{1}{\rho} \sum_{n=1}^{\rho} (i_n)_m$$

Discrete parameters need to be mapped back in discrete space, e.g. by probabilistic rounding. When the dominant (or discrete) recombination is performed, the offspring is a permutation of the parents features. Each parameter of the offspring k is a randomly chosen feature of one of its parents, $k_m \sim U(\{(i_1)_m, (i_2)_m, ..., (i_\rho)_m\})$. Same procedure is applied to the endogenous parameters assigned to the individual. The offsprings are generated until their number reaches λ. Note that an individual can be parent to many offsprings. There are multiple theories about how the recombination promotes optimality as well as diversity in this algorithm. For a detailed discussion of it see [15].

The most important tool to create variability in the population during the evolutionary algorithm is the mutation step. Again, the mutation procedure is applied to the endogenous

parameters and the parameters of the function under optimisation. Note however, that the endogenous parameters have to be mutated first as they affect the subsequent mutation of the function parameters. There are multiple requirements for the mutation process to make the algorithm effective. First, to ensure an effective search and enable a global convergence each point of fitness landscape and each parameter configuration must be reachable irrespective of the initial conditions and in a finite number of steps. Second, the mutation process have to be unbiased by fitness information. Selection and mutation serve antagonistic purposes, the recombination is a part of the latter process. Thus, it should not use the same underlying information as selection. Third, the entire evolutionary strategy system has to be defined in a way that enables a smooth evolutionary random path across the landscape. Hence, it is required that a small change in the parameter configuration of the function under optimisation results in a small change in the fitness. This depends not solely on the optimised or fitness function but on the parametrisation of the algorithm. Consequently, the optimised function is not necessary required to be differentiable. For more detailed analysis of the last requirement see [15, 98]. Nonetheless, the listed requirements can be violated without any loss of effectivity when the optimised function requires to do so and the violations are chosen carefully.

When a vector is mutated a randomly drawn offset is added to each entry of the parameter vector, $\mathbf{x}' = \mathbf{x} + \mathbf{z}$ where $\mathbf{x}'$ is the vector $\mathbf{x}$ after mutation. $\mathbf{z}$ is the mutation vector of length n with each entry drawn from a normal distribution and multiplied by σ_k, the mutation strength in a particular dimension. In mathematical terms, $z_k = \sigma_k Z; k \in \{1, 2, ..., n\}$, $Z \sim \mathcal{N}(0, 1)$. Thus, the strength of the deviation through mutation in each parameter of the optimised function can be varied. Note that the use varying mutation strength requires some knowledge about the system under optimisation. Consequently, the easiest assumption is $\sigma_k = \sigma$ for all k. In this case all mutants are normally distributed around the original point in parameter space in each dimension. While this procedure is suited for search in space of real numbers similar considerations can be done for discrete and binary domains [15].

After all offsprings were generated and mutated, selection is applied. While the previous steps increased the diversity of individuals in the population, selection is steering the system towards optimality. Depending on the implementation of the algorithm the next generation is composed exclusively from offsprings ((μ, λ) selection) or parents and offsprings ($(\mu + \lambda)$ selection). In both cases the μ fittest individuals are selected from the population to form the next generation of parents. In (μ, λ) selection scenario only the offsprings are considered for selection, when $(\mu + \lambda)$ selection is applied all individuals compete to enter the next generation. Thus, in $(\mu + \lambda)$ selection an individual can, in principle, survive over the entire algorithm duration. The selected μ individuals form the next generation. After new generation has been established the procedure to create the subsequent generation is repeated: parent selection, recombination and mutation, followed by selection. This iterative process is performed for intended number of generations. The last generation is a set of solutions which are optimal or close to optimal.

For many problems, the methods like gradient descent or simulated annealing are not feasible due to the complexity of the function, constraints, discrete and highly multidimensional parameter space. In these cases the evolutionary computations are the only family of methods to solve the task, combining the diversity of solutions achieved by random search and the optimality introduced by selection rounds. Note that the presented algorithm contains multiple parameters and methodic choices, which need to be made with regard to the problem at hand. The choice of mutation parameter σ can influence the width of distribution in the final generation. If σ is too large, the fitness of individuals in the final generation will fluctuate in a broad region around the optimum. Further under certain assumptions it can be shown that for real-valued search space the number of generations needed for the algorithm

to approach optimal solution scales linearly with number of parameters of the optimised function.

2.6 Existing approaches for resource allocation

The question of efficient resource allocation is of direct importance for public health. Major outbreaks can be prevented and the outcomes improved when resources are directed to the right locations efficiently and timely. This is a non trivial task as the spreading routes and the contact networks can be unclear, the delays unforeseeable. But even with knowledge of the underlying network and with negligible delays the task is rarely fully tractable, let alone analytically solvable. A similar problem arises in a different area where viruses spread, but not via human contacts. In computer networks the virus attacks or the cascading failure of nodes can be studied with the same methods as a contagion phenomena on networks in public health. Network structures in this field are usually known, top-down policies and compliance can be enforced easier than in human interaction networks. Public health and network security can benefit from sharing the insights, thus below we will discuss results from both fields, outline the differences and the application of results from computer science to public health.

All but one publication described in this section assume a top-down policy, which is enforced in a network by a certain overarching entity. This is a valid assumption, as there are global organisations capable of coordinating the international containment efforts. As will be outlined in chapter 6 there are multiple issues connected to such an arrangement. Further, all research papers discussed below consider two or more coupled locations which are affected or threatened by the disease. A simplest case to study is presented by two coupled nodes, connected by a transmission link. This topology is highly simplified and artificial, but analytically tractable. Thus it is often applied in combination with meta-population approach, meaning that the population inside the nodes is modelled by an ODE system. Most papers use the number of individuals infected in the course of an outbreak as a measure of the cost. The final outbreak size is acquired by explicitly solving the ODE equations or by an approximation. [93] and [64] used the described procedure to study the allocation problem with analytical tools. Epidemic models used are SIS and SIRS respectively. The later is also investigated in its limit case where the immunity vanes instantly, making it equivalent to an SIS model. In both publications all subpopulations are affected by the epidemic to different extent and the task is to distribute the resources to minimise the final size of the epidemic. While the ethics might suggest that the heavier affected region should be prioritised for treatment, both publications suggest otherwise. In fact, they show that treating the heavier affected region is the worst strategy among all considered alternatives. [93] concludes that the most efficient strategy is to concentrate the containment efforts on the less affected region and allocate no resources to the heavily affected region. While ethically questionable, this strategy results in much lower global prevalence.

In the paper by [93] resources are allocated prior to the epidemic as a one time donation. [64], on the other hand, proposes a model in which a certain constant amount of resources is supplied per time step. Thus, a more complex strategy can be employed. The optimal strategy presented in the paper requires to change the recipient of investment at a certain time point. Most affected region receives investment at the early stage, from a certain time point onwards the resources are committed lesser affected region. The timing of the switch is highly dependent on the disease specific parameters, which makes this strategy difficult to implement in a real world scenario. [116] has shown that it is possible to solve a more complicated system of connected populations.

When the problem is studied on a more local scale, the nodes are assumed to be individuals, thus assigned a discrete state. This approach is also pursued when computer networks are studied. This way the efficiency of a containment strategy can be evaluated with respect to a specific population structure. Many papers have studied the targeted vaccination without resource constraints as has been outlined in section 2.3. Several publications, published in computational journals, propose solutions to this problem under constraints, e.g. [87]. In this paper resources of two kinds are available: preventive, vaccine like resources and corrective antidotes, which can be given to an affected node to free it from the disease. The paper proposes an algorithm to solve the resource allocation problem using geometrical programming. The final solution resulting from an extensive calculation, does not correlate to any of the existing centrality metrics and results in a complicated pattern. Preciado et. al. apply proposed algorithm to computer and transportation networks alike. [20] studies the problem in a more analytical manner, correlating the final solution to existing metric. His research indicates that the allocation strategy can be guided by degree of the nodes, but decision whether high or low degree nodes are granted investment depends on the epidemic regime of the system.

The resource allocation in context of multi-resistant bacteria spreading over a hospital network was studied by [85]. In the proposed model allocating resources to a hospital decreased the probability of disease transmission to the respective node, thus slowing the overall disease progression. In this paper heuristic optimisation algorithms are proposed and implemented, which are capable of calculating an effective resource distribution over the network. The paper demonstrates that the algorithm can calculate resource distribution which is three to six times more effective than the strategy currently employed by the hospitals. The algorithm can be used outside of the public health domain as long as the spreading takes place on a network with known topology and at the time of deployment infected nodes are known. The latter must not be true in a realistic situation due to gaps in surveillance or diseases where asymptomatic progression is possible. [118] proposes multiple algorithms to calculate optimal resource allocation given unreliable and incomplete information.

The only paper to our knowledge, which studies a bottom-up, agent driven resource allocation is [103]. The system under investigation consists of two connected meta-populations, which own a specific amount of resources and can distribute available resources to remove susceptible from the population, e.g. in analogy to vaccination. In this paper the nodes are required to spend all available resources on disease prevention, but can choose how to distribute those between the nodes. The distribution of resources is chosen to reduce the final number of affected individuals among own meta-population. For each node there is an incentive to donate a fraction of resources to the second subpopulation as it poses a threat to the local population through the coupling term. Wang et al. prove that in this system a Pareto optimal solution exist. They further study the outcome of a top-down, global optimisation solution. In this case overall number of infected is being minimised. As expected from similar problems studied in context of game theory, there is a mismatch between selfish optimal solution and the global optimisation. When optimised selfishly the infected neighbour receives less investment than is considered optimal from the global perspective. Note that this model is not a subtype of a public good game, but only resembles it in minor details.

2.7 Summary

As has been outlined above, the epidemiology is a rich field being explored by a variety of methods. After the introduction of network approach into epidemiological field modelling global spread became feasible and common in times of ongoing pandemics. Aside from

forecasting the progression of a certain outbreak in real time, efforts have been made to draw more general conclusions, which can be readily used without reliance on knowledge of disease specifics. Latter approaches, on par with sophisticated modelling tools, can inform decision makers about future development of an epidemic to enable a timely and adequate reactions.

Further research has been made to understand what an adequate and feasible reaction is. The game theory has been a tool of choice to study resource allocation as soon as conflicting interests arise. One of the current topics widely explored using this approach is the vaccination and herd immunity. Multiple relevant insights were generated, among others the impossibility of achieving herd immunity when all individuals act perfectly selfish and the emergence of fluctuations in the vaccine uptake. The resource allocation on a global scale with response to a pandemic was investigated for a set of simplified systems. The research in this area was concentrated on a top-down enforced allocation of available goods among all actors of a system. It has shown that optimal strategies can generate ethical conflicts, i.e. when the optimal strategy prescribes to abandon heavily affected countries and aid the regions with fewer cases. Furthermore some more complicated strategies were proposed, but their reliance on exact parameter estimates makes them a risky choice in a real world scenario, where disease parameters are difficult to estimate and are subject to seasonal and regional fluctuations.

The work presented in this thesis is at the crossing of two approaches described above. On the one hand it aims to develop a set of methods to evaluate the risk of disease importation and identify influential nodes. On the other hand it describes a solution to the resource allocation problem on a global scale, but releasing the top-down assumption. Assuming selfish interests, a conflict among the players arises, making the game theoretic optimisation a natural choice. To our knowledge there is only one publication studying resource allocation by selfish agents. It demonstrates that optimal allocations are different among players, highlighting the conflict. Further, when optimisation minimising global cost is performed the outcome differs from the selfish optimisation. As will be demonstrated in later chapters, this result is consistent with our findings despite the different methodology.

Chapter 3

Prerequisites

Research presented in this short chapter is necessary to understand the methods introduced in chapters 4, 5 and 6. Unlike the broad overview given in chapter 2, following pages give a detailed description of two Concepts. First we construct and characterise the world aviation network, which serves as a basic structure for the metrics developed throughout this thesis. Second, we describe the *effective distance* introduced by [17] and used extensively in chapters 4 and 5. Forthermore,

In the work at hand we make use of the concept of effective distance introduced in paper [17]. This concept was derived for transportation networks where the fluxes between the nodes are known. Its use originally demonstrated on the world air transport network (WAN), the same network we will use in this thesis. Multiple publications has highlighted WAN as the basis for global pandemic spreading [25, 41, 67, 101]. It has been shown that predictive power of pandemic models is increased when world air traffic is considered, as it is the fastest way of long distance travel with million passengers per day [58]. Due to short travel times an infected is unlikely to recover on a plane, thus carrying the disease to his traffic destination. This makes air traffic an important transmission path. Very short travel times are an important change compared to the traffic about hundred years ago, where journey between continents took months to finish. Back then an infected was likely to die or recover prior to his or her arrival. This is one of many reasons why modern pandemic spread with a much greater speed than their historic counterparts. Traffic volume is another reason for increased spreading speed. Effective distance and effective distance trees can be derived from any kind of transportation or freight networks. In the original paper the methodology was successfully applied on food supply network in Germany to track food bourn outbreaks.

In the following chapter we will explain effective distance and introduce the world air transport network. First we will we describe the WAN in greater detail, data which served as a basis and the construction of the network from traffic data. Further we analyse the distribution of traffic across continents and regions, discuss how the amount of traffic in a region relates to its population and the implications of this relationship. Second we demonstrate how effective distance can be calculated on WAN, how it can be used to predict the course or the source of a pandemic and to estimate arrival times of the disease.

3.1 World air transport network

World air transport network (WAN) is derived from the data of scheduled flights as reported by [58]. Data contains 3856 airports and 26691 flight connections. The network is constructed using airports as nodes and flight connections as links. The weight of a link from node i to node j, F_{ji}, is defined as number of seats on scheduled flights from i to j during the year 2014[1]. We further define relative flux of a node i, $f_i = \sum_j F_{ij} / \sum_{k,m} F_{km}$. Relative flux over a node represents the fraction of worldwide traffic accumulated by node i. Note that $F_{ij} = F_{ji}$ as the network is symmetrical. Furthermore geographical coordinates, region and country to which the airport belongs is provided by the dataset. Hence, we use this information throughout the thesis.

The WAN is highly inhomogeneous network with multiple high degree hubs and a much larger number of low degree nodes. It shows distinct small world properties. Median and mean relative fluxes are $\tilde{f} = 1.6e-5$ and $\bar{f} = 26e-5$. In this network the top 100 airports account for more than a hsalf of the entire passenger flux. The diameter of the network is 16. Although, when smallest airports accounting for 2% of the global flux and links connected to those airports are removed, the diameter is reduced to 6. Reduced network contains 1787 airports, approximately a third of the nodes from full WAN. This underlines that high fraction of nodes present in the WAN accounts for a little fraction of links, while the core of the network is composed from a smaller number of well connected hubs.

The WAN gives insight about the distribution of the air traffic across the globe and can give us a general understanding of the participation of different nations in the air transportation. While we generally expect the flux to be roughly proportional to the population inside of the catchment area, statistics provided by the World Bank allow us to verify this assumption at least on the level of countries, regions or continents [10]. The population percentages represent the fraction between the population in the respective location divided by the total worldwide population. Flux fractions are calculated analogously.

On continent-wide scale the proportionality between population and flux is weak. While Asia is home to more than 60 % of the global population, it accounts for about 37% of the air traffic. Similar situation arises in Africa: while it harbours approximately 15 % of the global population, its share of the air traffic is 3.2 %. It also hosts 353 airport, which is the second lowest number in the WAN. Oceania is the only continent accounting for less airports than Africa, home to 337 airports. At the same time, Oceania's population is only 3.5 % of the population on African continent. With exception of Africa, ranking of continents according to population and to flux match. Hence we can conclude that while there is a proportionality between population size and amount of flux, it is neither direct and nor linear. When assuming equal proportionality across the globe, systematic under- or overestimation can lead to false conclusions.

When populations and fluxes on regional level are considered, similar picture arises. Generally, higher flux in a region correlates with bigger population, nonetheless several regions deviate from this rule. Regions with better developed infrastructure account for higher fraction of flux, e.g North Africa is accounting for 36% of african traffic while it harbours 21 % of its population. Middle Africa, on the other hand, accounts for only 6 % of African traffic and 13.5 % of the population. In Asian region, Central and Southern Asia participate less in the air transportation than expected from their population fractions. While Southern Asia harbours 42 % of the Asian population, it accounts for only 10 % of the traffic. Western Asia,

[1] Raw link weight matrix $\mathbf{F}_{raw}$ represents cumulative number of seats in scheduled commercial flights between target and destination over the year. To ensure that $F_{ij} = F_{ji}$ at every link, the flux matrix $\mathbf{F}$ was defined as $\mathbf{F} = \mathbf{F}_{raw} \times \mathbf{F}_{raw}^T$. Division by 2 was omitted as the absolute values of passenger flux are not important for the methodology presented in the work at hand.

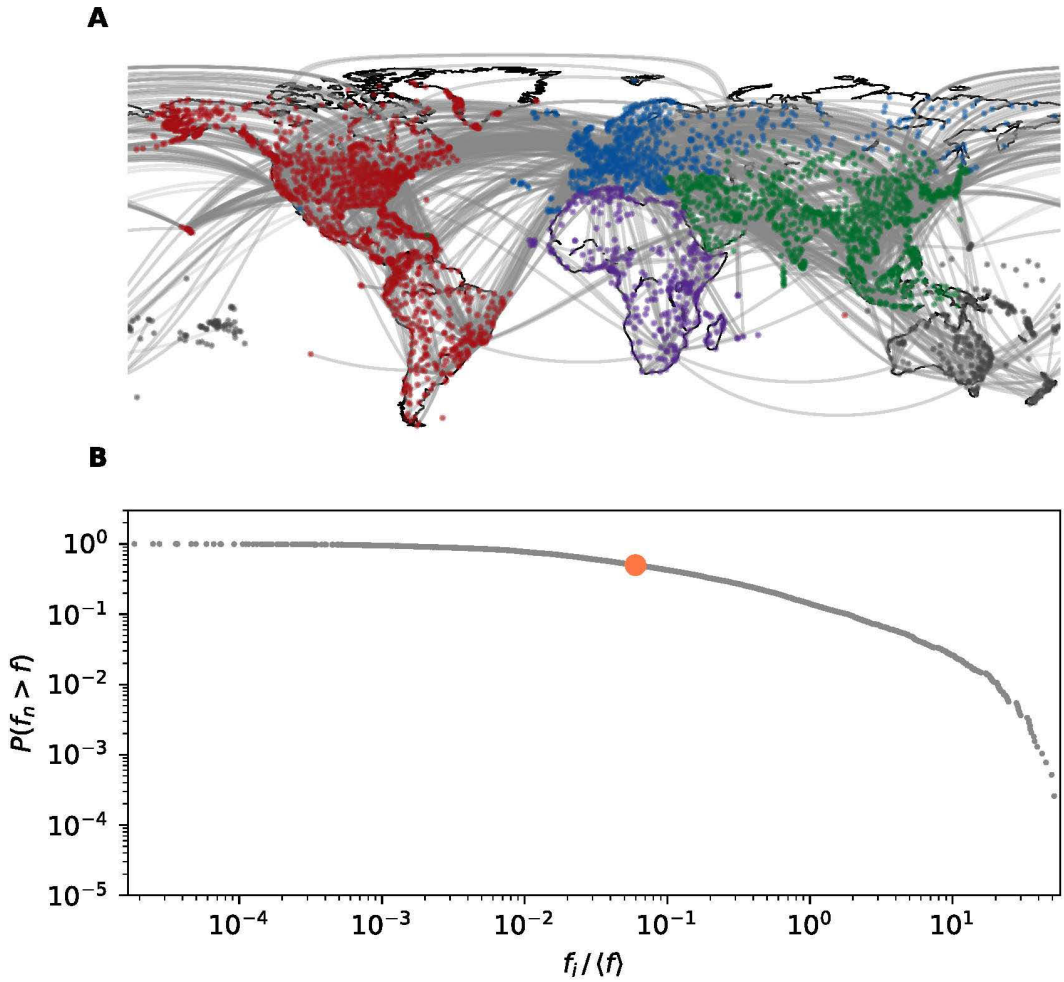

Figure 3.1: (**A**) The structure of the WAN superimposed with a geographical map; points show the location of the airports, lines show flight connections, colors represent the continent. The graph shows traffic connections with highest passenger numbers, presented links account for 50% of the worldwide air traffic. (**B**) Relative flux distribution of the WAN. X-axis shows the relative flux accumulated by a node, y-axis - the probability that a randomly chosen node has relative flux higher than the x value. There is a number of hubs accounting for more than one percent of the global traffic each. It also contains a high number of nodes which contribute little to the worldwide traffic flow, mean and median weighted degree are one order of magnitude apart. The WAN is a highly heterogeneous network showing small world properties.

on contrary, is the smallest Asian region with 6 % of the population. Nevertheless, it facilitates 16 % of Asian traffic. On American continent, North America accounts for 73% of the entire traffic, bearing only 36 % of the population. American continent as a whole hosts 1514 airports, which is the highest number of airports per continent. 990 of American airports are situated in North America. Flux distribution across Europe is more even, with Eastern Europe accounting for smaller and North Europe higher fraction of the air traffic, compared to the expectation from the population of the region.

The distribution of the hub airports across the world is uneven, with most hubs being located inside the developed countries. 30 out of 100 biggest hubs are part of the North American region, 20 are located in Eastern Asia and 13 in Western Europe. Central America and South Africa are home to one of the top 100 hubs respectively. Other regions represented by the top 100 hubs are multiple European regions, Australia, South America, Southern and South-Eastern Asia.

The above observations emphasise that on par with the amount of population, developmental level of the region impacts the flux inside this region. Regions harbouring the highly developed, high income countries account for higher fractions of air traffic and host more global hubs. As the hubs are used as the transit nodes, this will have implications in the latter chapters. In general, it is likely that an agent starting at a random node will traverse one of the hubs after only a few steps. This means that highly developed more probably host transit nodes and lie at short effective distance to the rest of the network.

3.2 Effective distance

The main idea behind the effective distance is that the spreading on a network is dominated by the most probable paths, leaving the rest of connections redundant. This is analogous to the dominance of the smallest resistor in an electrical circuit, where many heterogeneous resistors are connected in parallel. This allows us to simplify a complex network to a tree, if we know the outbreak origin. This method was originally described in [17].

Consider a weighted network with N nodes and L links. A link from node i to node j has a weight of w_{ij}. Some special attention has to be payed to the definition of a weight. One possibility to define a link weight is to consider it to be the cost connected with traversing an edge (analogous to resistance) or as distance along this edge (analogous to road distance). In this case a higher link weight means longer distance between the nodes. Another approach is to define a link weight as flux over an edge (analogous to the amount of passengers travelling along a route), relating bigger link weight to shorter distance between the nodes. In the thesis the latter definition is used, in case of WAN $w_{ji} = f_{ji}$.

When a process is spreading over a network it can be carried by a random walker, in analogy to an infected individual travelling via air traffic. To predict the course of the spreading, we need to know the route of the walker. At each node i the walker has a limited set of links which he can take. To be precise, nodes which are not connected to i via a link can not be reached by the walker in a single step. Hence, the probability that at the next step the agent will move to one of i's neighbours is $\sum_{j \in \mathcal{N}} P_{ji} = 1$, where $\mathcal{N}$ is the set containing all neighbours of i and P_{ij} is the probability that a walker will travel from i to j. The definition

$$P_{ji} = F_{ji} / \sum_k F_{ki} \tag{3.1}$$

satisfies this condition. Note that even in case of $F_{ij} = F_{ji}$, $P_{ij} \neq P_{ji}$ in most of the cases. When speaking about distances, we expect them to be additive, hence we define the effective length from i to its neighbour j as

$$d_{ji} = d_0 - log P_{ji}$$

This way effective lengths along one path can be added and lower probability P_{ji} results in a bigger distance between i and j. In addition we can define a minimal distance between two nodes, d_0. If we wish to penalise the paths involving multiple transit nodes, we can assign $d_0 > 0$. We can think of this penalty as a waiting time until the walker can move on to the

next node. Throughout the course of the thesis we define $d_0 = 0$. All presented results were also tested using $d_0 = 1$ without any qualitative and only minor quantitative changes.

To calculate the overall length of a path from n_0 to a remote node n_L in terms of effective distance we need to sum the effective lengths of links between the nodes contributing to this path. Using above definitions we can calculate the length of a path from n_0 to a remote node n_L along a path $\omega = \{n_0..., n_L\}$

$$\lambda = \sum_{i=1}^{L} d_{n_i n_{i-1}}$$

and define effective distance between the nodes as the shortest length of all possible paths

$$D_{n_L n_0} = min(\lambda)$$

Using the shortest paths based on effective distance we can construct the effective distance tree rooted at node i. This tree shows all most probable spreading routes from i to the rest of the network. Note that on the effective distance trees, just as on the shortest path trees, issues arise when multiple shortest paths between a pair of nodes exist. This is not the case in the WAN and is generally an exception in weighted, non highly symmetrical networks.

Modern epidemics result in wave-like patterns when the nodes are remapped according to the effective distance to the origin of the outbreak. This highlights that the spreading in the modern world is not dominated by the geographical distance, but rather by the possible fastest routes of travel. For individual, international traffic the air transport is unprecedented in its speed and volume of traffic. Hence, it is not surprising that it determines the international spread of diseases. Nonetheless, usefulness of the WAN can decline when short distances are concerned. It is likely that commuter and short range travel is dominated by car or train traffic. When investigating such scenario we advice to use a different network, which accounts for most used means of travel in context of the problem.

An important application of the effective distance tree lies in the estimation of the arrival time of the epidemic. The latter can be calculated without the knowledge of any disease specific parameters, which makes this approach more robust against the uncertainties in the estimation. The arrival time can be described as the quotient of effective distance from the source over the spreading speed of the disease on the effective distance tree

$$T_a = D_{ji}/v_{eff}(\alpha, R_0, \gamma, \varepsilon)$$

The spreading speed can be estimated as soon as the disease has been transferred to at least one node different from source. Spreading speed depends on disease specific parameters, but knowledge of these is not required. Knowing the effective distance from the source to this node, D_{ki}, and the time difference between the beginning of the outbreak and its arrival at node k, Δt_{ki}, we can estimate the spreading speed as

$$v_{eff} = D_{ki}/\Delta t_{ki}$$

As has been demonstrated in the original paper, the effective distance can also be used to estimate the outbreak origin, in case it is unknown. As highlighted by the formula, there is a linear dependency between effective distance and the arrival time. Hence, by calculating the correlation between these values most probable outbreak origin can be identified. For a more detailed explanation and application examples refer to [17].

Throughout this thesis we will use the effective distance and the derived tree to define further metrics which we apply to guide decision making during an unfolding pandemic.

Chapter 4

Defining the scope: A context specific approach to identifying key airports during a pandemic

In this chapter we introduce two metrics: *scope* and *confluence*. Both metrics describe the importance of a node during a pandemic and its role in dessiminating the disease. Defined metrics rely on effective distance trees and therefore present a context sensitive characterisation of nodes in a specific outbreak scenario. After showing the mathematic definition of the metrics we superimpose them with established, non-context sensitive centrality measures. On the world air traffic network, scope and confluence result in a more diverse and specific list of super-spreader nodes across multiple outbreak scenarios. Further, we demonstrate how airport profiles can be defined in terms of scope and be utilized to group the locations according to their region of influence and the characteristic role they take on. We show that the airports can be separated in two groups: *generalists* and *specialists*. Specialists play a major role during outbreaks in a specific region while having a minor influence when the outbreak is initiated outside of this geographic region. Generalists, on the other hand, are important disseminating points in many outbreak scenarios. Finally, this chapter shows how the matrics can be applied to a selection of hypothetical and real world outbreaks. Moreover, the metrics can be applied with the focus on a target location, informing in what scenarios this location takes on an important role.

On early stages of epidemiological modelling, ODE systems have dominated the field. At the heart of this approach is the assumption that all agents in the system are equal, surrounded by the same environment and have the same probability of contact with every agent in the system. This assumption has big implications for the outcome of the model and hence introduces a bias. Recent research has shown that in real systems this assumption doesn't hold. A more realistic way of describing social systems was offered by network theory, where individuals are approximated by nodes and relations between them are described in terms of links. Incorporating this approach made it possible to account for differences in the number and strength of contacts, removing the uniformity assumption. This also changed the way how countermeasure deployment was viewed. When all agents are considered to be equal and live in the same environment, there is no difference in whom to target by the interventions, e.g. vaccination or quarantine. On networks some nodes can be a better target for interventions than other, e.g. because of a high number of links or their central posi-

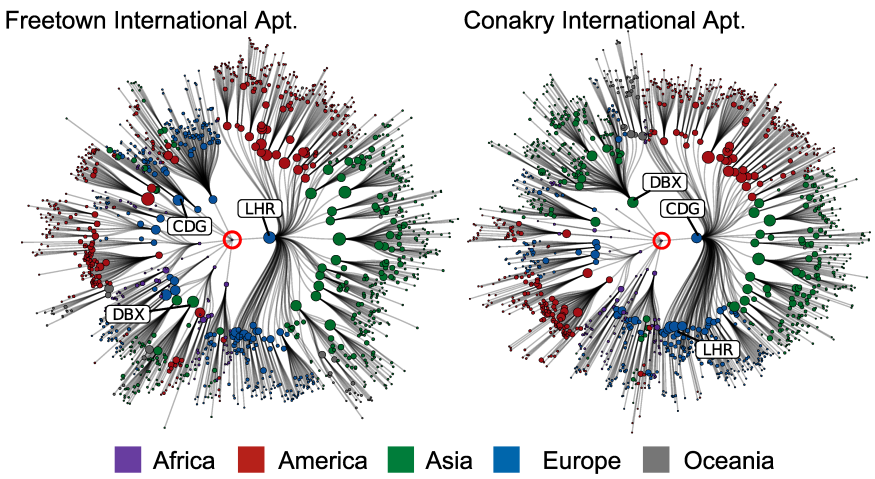

Figure 4.1: Examples of shortest path trees derived from air transportation network as explained in [17]. Outbreak origins are located in Freetown (left) and Conakry (right), the length of the branches represents the effective distance from the origin. Nodes are coloured according to the geographical region. While the trees look similar in their structure different nodes take on important roles. London Heathrow (LHR) is an important gate in case of an outbreak in Freetown, but it has only a minor role when the epidemic starts in Conakry. Paris Charles de Gaule (CDG) displays the opposite behaviour. Dubai International (DXB) plays a much bigger role if the outbreak is starting in Conakry compared to Freetown. Therefore the knowledge of the origin and the structure of the shortest path tree is crucial for effective countermeasure deployment.

tion between communities. There have been efforts to use conventional centrality measures from graph theory to guide decisions about intervention efforts on networks [99, 51, 39, 28]. While there is a myriad of metrics, some of them have gained bigger popularity than others [70]. The simplest centrality measure is degree centrality, which describes the number of neighbours of a node. This metric indicates how many nodes are directly influenced by the source or can be infected by it. Another metric, which was was demonstrated to be useful in many respects, is betweenness centrality. This centrality measure accounts for the number of shortest paths crossing a node. This reflects how many long distance connections are disrupted when a node is removed from the network. Other common metrics are closeness centrality, which reflects the shortest path length from a node to all remaining nodes of the network, and eigenvector centrality, which considers not only own degree, but also factors in the degree of surrounding nodes. All centrality measures mentioned above were developed outside of the epidemic context and are calculated for the network regardless of its state, not including any information about the source of a spreading phenomenon. Thus they assign a node the same importance in every pandemic scenario.

Figure 4.1 demonstrates that such thinking obscures the real picture of the highly variable role of a node depending on the epidemic scenario. The graphs show effective distance trees, as explained on page 30, rooted at Freetown, Sierra Leone, and Conakry, Guinea. Both cities were heavily affected by Ebola epidemic in 2014 and posed a potential threat of exporting cases. Their locations are geographically close, lying only 217 km apart, but the route via which the disease would most likely spread globally from either source is very different. The graph demonstrates that Paris Charles de Gaule airport (CDG) is a major transit airport in case of an outbreak in Conakry. This means that the probability of an infected who exits

Conakry to pass through CDG, weather to exit there or to proceed travelling to another destination, is very high. If the outbreak happens in Freetown, Charles de Gaule airport has much less importance. The inverse is true for London-Heathrow (LHR) airport, which is a major transit airport when the journey is started in Freetown, but is close to being a leaf in a tree rooted at Conakry airport. This highlights the need of a context sensitive metric which accounts for the origin of the outbreak.

Contrary to the above, different conventional metrics often favour the same hub nodes as we demonstrate in figure 4.2. The figure shows the top 15 airports from the WAN according to passenger flux, which is equivalent to weighted degree, betweenness and closeness centrality. As demonstrated by the plot, betweenness and closeness centrality correlate heavily with weighted degree. If each metric would provide an entirely distinct ranking, figure 4.2 would contain a set of 45 unique airports. Instead, only 27 distinct airports are present in the figure. Furthermore, 24 airports are present in at least two centrality measure rankings, 4 are shared between all three. This highlights how much overlap is observed between these metrics. Considering the computational power needed to calculate complex metrics like betweenness and closeness centralities and the ranking overlap with flux or degree centrality, the benefit from the calculation might be minor. None of these metrics is solving the problem of a varying node role depending on the origin of the spreading phenomenon. Hence we need to develop a new, context sensitive metric which will account for the variability observed in areal world scenario.

In this chapter we approach this issue by defining new centrality measures following the methodology introduced in [17]. The outbreak origin enters the calculation, thus we account for the specific context in which our metrics are to be applied. We use shortest path trees based on effective distance to derive two node properties: *scope and confluence*. We see the latter as a context specific number of children to which the disease can spread directly through the node. Scope accounts for the fraction of the network which can be accessed on the shortest path tree through the node. This also reflects the fraction of flights which will use the node either for transit or as final destination. Hence by deploying interventions at this node we can protect a given fraction of the population. We demonstrate how those metrics relate to existing centrality measures, particularly degree and betweenness centrality, and show that the later is a special case of scope. We outline how a set of trees can be used to calculate average scope or confluence if a wide geographical region is affected by the disease or the true outbreak origin is unknown. Averaging has some implications which will be discussed in this chapter. We demonstrate that scope is more suited to identify nodes of importance for a particular outbreak scenario than established metrics by being context sensitive and giving a tractable value for the benefit connected to the deployment of countermeasures in a particular node. We apply scope on different outbreak scenarios and discuss which nodes act as transit locations and how influences of colonial history can still be observed in the structure of world aviation network (WAN). One of the advantages of scope is that it is possible to switch perspectives and concentrate on a target node rather than the outbreak origin. We show how outbreak sources can be classified according to the scope value they invoke for a specific node. This application is important for local authorities who monitor and manage a certain airport. Thus they are in a position to prevent disease spreading beyond the respective airport if the response is appropriate.

4.1 Definition of scope and confluence

As has been demonstrated by [17], spreading processes on networks are dominated by most probable paths between nodes. Hence we expect that the import of disease to a new node

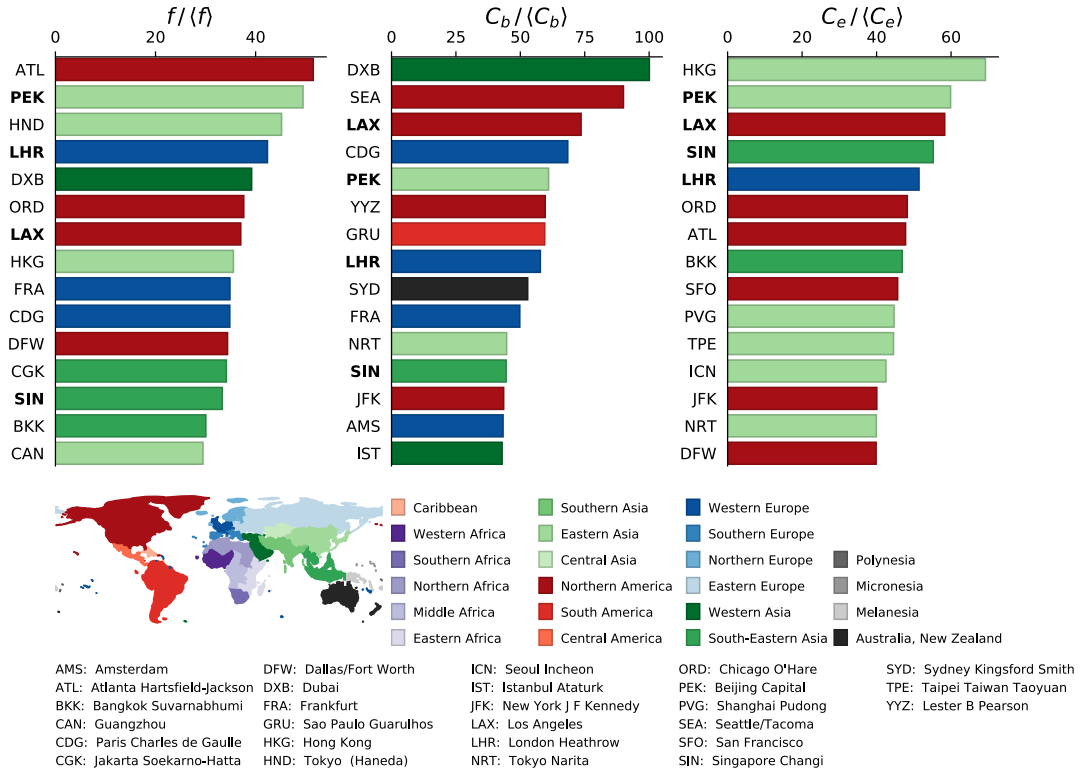

Figure 4.2: Top 15 airports according to passenger flux f, betweenness centrality C_b and closeness centrality C_c. Airports are labeled by respective IATA letter code. Bold labeled airports are part of the top 15 according to all three centrality measures. Colour of the bars indicates their geographic location. Established centrality measures favour similar nodes. Out of 27 nodes defined by the top 15 among all centrality measures, 24 are shared in at least two out of three metrics.

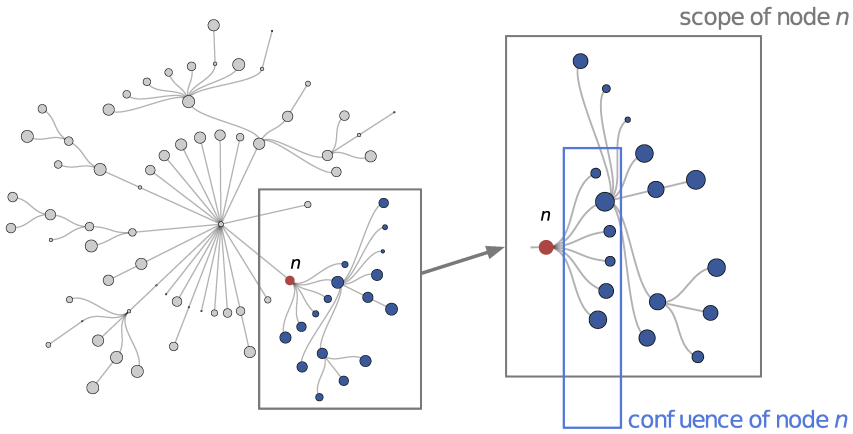

Figure 4.3: Graphic explanation of scope and confluence. The shortest path tree is derived from a random geometric graph, an example node n is shown in red. Confluence of node n is the number of paths, which converge at the node. Nodes with high confluence present good intervention targets as each node downstream from them will cover only a fraction of paths joining at node n. The fraction of the network population on the branch downstream from node n, itself included, is the scope of the node (blue and red). Nodes with high scope act as transit bottleneck to a big portion of the network in case of an epidemic and represent valuable targets for interventions like passenger screening.

will happen via such probable paths and less probable paths can be neglected. Note that in some networks there can be multiple paths with the same probability. In this case neglecting either can lead to biased results. However, only highly symmetrical graphs, like lattices, are affected by this problem. The the WAN, which is highly heterogeneous, has a very low chance of multiple equiprobable paths. Further in none of the presented cases did we observe multiple paths with same or similar probabilities. Therefore we define new metrics, scope and confluence, based on the effective distance tree instead of a full network. As a consequence both metrics depend on the origin of the tree. Below we will demonstrate that scope and confluence can vary greatly depending on the outbreak scenario. We will show how those metrics can be averaged over multiple trees and how this changes the reliability.

Given a network and respective effective distance tree T_i with node i as its origin, we define confluence of the node $c_n(T_i)$ as the number of its direct offsprings on the shortest path tree. Confluence of a node is strictly $k_n - 1$ or smaller, where k_n is degree of node n. Scope of a node s_n is defined as the fraction of network population which can be reached through node n on the shortest path three

$$s_n(T_i) = \sum_{k \in \Theta} \eta_k; s_n(T_i) \leq 1$$

where η_k is the population fraction at node k and $\Theta = \Theta(T_i)$ is a set of children, grandchildren etc. of node n as well as n itself. Thus, the most probable path to each node in the set Θ is traversing node n. Note that Θ is dependent on the effective distance tree and hence the outbreak origin. For the tree origin i this set contains all nodes of the network, hence $s_i = 1$.

A note regarding the population size of a node is necessary. The most natural approach to define the size of a population of the airport node in the WAN is by assigning it the size of the population in its catchment area. This information is rarely available. Thus, one

possible assumption is that the population size inside the catchment area is proportional to the flux. This assumption is only true for the mobile population inside the area, meaning the population fraction which has real access to the air traffic. Due to low income, availability of faster transport or geographical boundaries some population fraction can be excluded from the access to the airport. Assuming the aforementioned proportionality of flux and population inside the catchment area, systematic errors can arise. For example we are given two nodes, n and k, and $f_n = f_k$. By proportionality assumption we set $\eta_n = \eta_k$. If, due to social disparities or any other reasons, a higher fraction of n's population has access to the air traffic than at node k even if the population at n is smaller it will generate the same flux as the less mobile population at k. This way we underestimate the value gained by protecting the population at node k.

Thus scope can systematically overestimate or underestimate the importance of some nodes. For this reason we encourage the use of real population data if available. To demonstrate the methodology we will estimate the size of the mobile population from available fluxes. We thus assume that mobile population N_n at node n is proportional to the flux of the same node F_n. Hence the same must be true for the relative values $\eta_n \sim f_n$. The most general proportionality is $\eta_n = \alpha f_n^{\varepsilon} + \beta$, where α, β and ε are factors. For the sake of simplicity we use direct linear proportionality $\eta_n = f_n$ for all nodes.

Our approach until now assumes that the true origin of the outbreak and the tree is known, but there are multiple scenarios where this may not be the case. On the one hand, we can think of a scenario where the outbreak occurs in multiple, possibly neighbouring locations. This will generally happen when the outbreak is not contained fast enough and spreads via the air transportation route as well as locally by means of other traffic. On the other hand, true source of a disease may not be known at the beginning of a pandemic, e.g. when cases are reported in multiple countries in fast succession. One of the most recent examples is the N1H1 pandemic of 2009, where the origin remains unclear even multiple years after the pandemic has passed. We can account for both situations by calculating average scope over multiple trees. Thus if we have a set of possible outbreak locations $Y = \{i_1 ..., i_L\}$. We can average the scope such that

$$s_n(Y) = \sum_{i \in Y} \alpha_i s_n(T_i) \tag{4.1}$$

where α is a weighing coefficient. One possible definition is $\alpha_i = 1/|Y|$ for all i, resulting in an equally weighted average. Nonetheless, it is plausible to assume that an outbreak in a bigger meta-population will pose a higher threat compared to a smaller one, as the disease can infect higher absolute number of people. Thus we expect a big meta-population to export higher absolute number of cases and want to reflect that dependency in the average. Therefore in this thesis we define $\alpha_i = \eta_i / \sum_{k \in Y} \eta_k$. The same procedure can be followed to average confluence.

Both metrics can vary greatly for a node when the trees are rooted at different origins as demonstrated in figure 4.1 for LHR and CDG. Therefore averaging must be applied with reason. In some realistic scenarios it can be more informative to inspect scope resulting in each scenario separately, instead of taking the average. For example when deploying passenger screening it is important to know which outbreak location poses the highest threat and concentrate the efforts on incoming flights from it. Nonetheless the averaging method can be used if little information about the outbreak is available or if a bigger region is affected. Furthermore scope and confluence averaged across all possible outbreak origins can be compared to existing centrality measures.

Betweenness centrality C_b and scope s_n are based on a similar reasoning: nodes, which

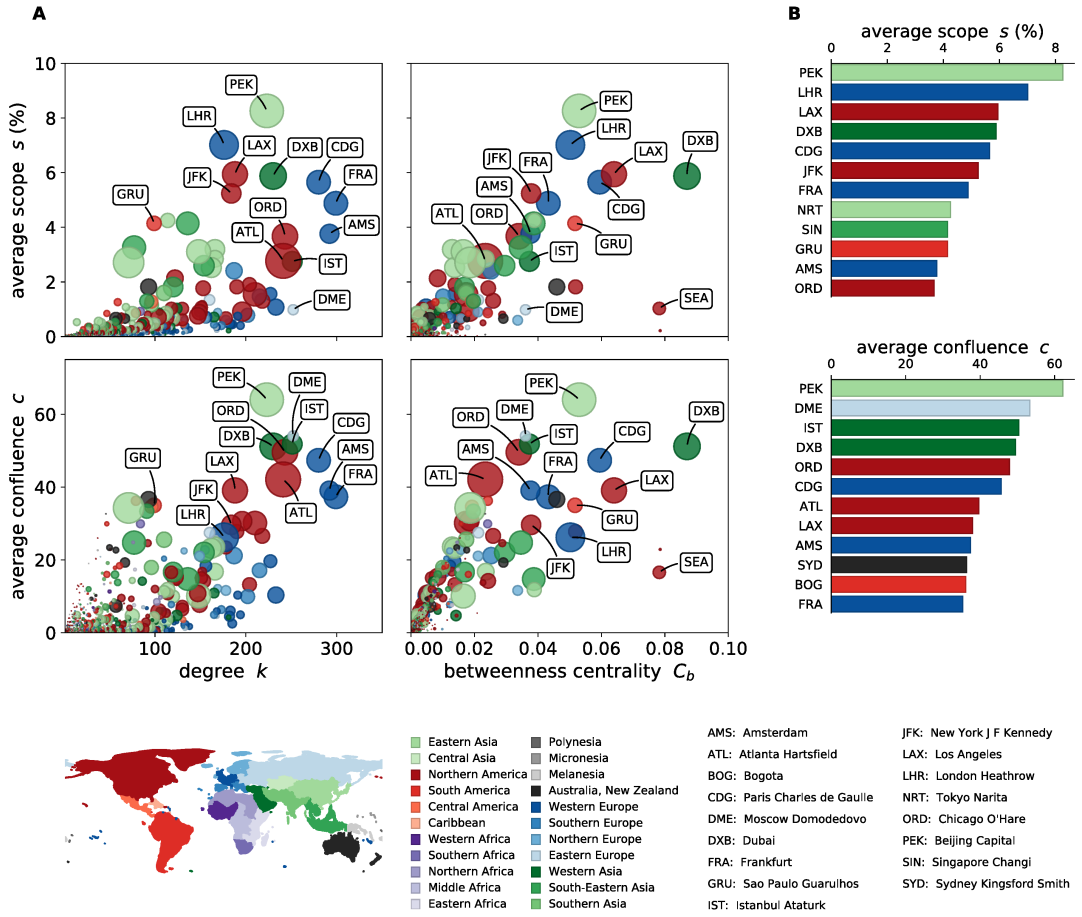

Figure 4.4: Relationship between proposed and established centrality measures. Colour code is according to the region of the airport. (**A**) Established centrality measures versus average scope and confluence. The size of the nodes is scaled with passenger flux. Selected hub airports are labeled with the IATA codes. All metrics show a similar trend, but none can be used as a good predictor for the others. Beijing International Apt. has the highest scope and confluence while its degree and betweenness centrality are considerably lower, possibly understating its importance during epidemics. The reverse is the case with Seattle Apt. which shows high betweenness centrality and very low confluence and scope. (**B**) Top 12 airports according to average scope and confluence. Only 6 airports are shared between the bar charts There is considerable overlap with the top 15 airports reported by other centrality measures (Fig 4.2). All 4 airports shared by 3 conventional metrics also appear in scope ranking in the top 12, further 6 airports are shared with at least two other centrality measures. Confluence deviates more, sharing only 7 airports with two or more conventional metrics.

constitute to multiple shortest paths are important. We can show that under special conditions betweenness centrality is a special case of scope, provided path length is judged based on effective distance in both cases. On a regular unweighted graph (i.e. $F_{ij} \in \{1, 0\}$ for all i, j) with $Y = \{0, 1, ..., N\}$ scope and betweenness centrality have a linear relation of a form $C_b(n) = s_n(Y)/f^*$ with f^* being the relative flux per node. For detailed derivation see Appendix 8.1. Note that betweenness centrality always considers paths from all origins, while scope can be limited to one or few trees. This way scope can be used in a context sensitive manner providing more specific information for the scenario at hand, while betweenness centrality can only provide the global average.

Figure 4.4 **A** compares average scope and confluence with unweighted degree and weighted betweenness centrality. We have chosen degree as the second conventional centrality measure as its logic is most closely related to the concept of confluence. While all metrics have a similar trend many nodes are assigned a different importance according to different metrics. For example Beijing Intl. airport (PEK) is ranked as the top airport according to average scope and confluence while its degree and betweenness centrality are not the highest. A similar picture arises for London-Heathrow airport (LHR), which role is far less according to betweenness centrality, while its scope ranking is second highest. The most striking example is Seatle-Tacoma airport (SEA), which shows the second highest betweenness centrality, but scores very low in terms of scope and confluence. Figure 4.4 **B** shows the top 12 airports according to average scope and confluence. There is still a considerable overlap with conventional centrality measures, but multiple new airports appear in the ranking. PEK and LHR are not ranked as top airports by any of the conventional metrics, but are being close to the top according to scope. Nonetheless high overlap between average scope and other centrality measures indicates that scope averaged over all possible locations suffers from the same drawback: much context specific information is lost. Hence the advantage of scope and confluence is negated by global average.

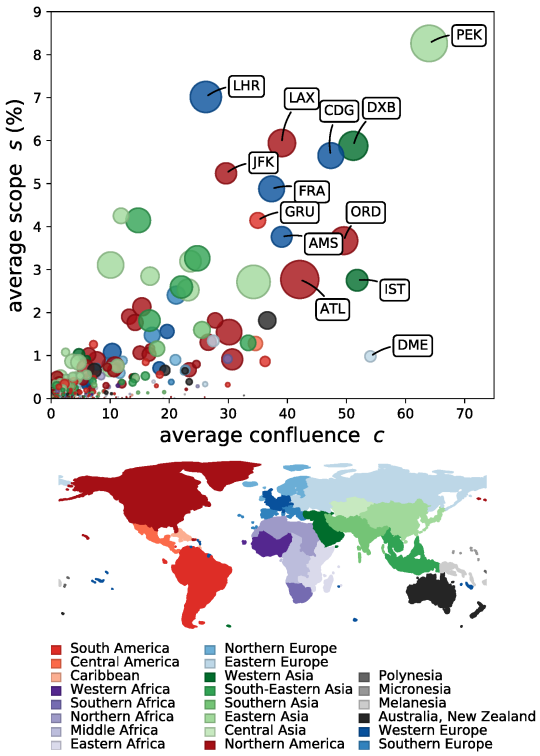

Figure 4.5: Relationship between average scope and confluence. Nodes are scaled with the passenger flux and colored according to the geographical region. Both metrics show a similar trend. There are several exceptions showing low confluence and high scope (London Heathrow, New York J.F. Kennedy and Los Angeles International) and vice versa (Hartsfield–Jackson Atlanta, Istanbul Atatürk and Chicago O'Hare).

One important phenomenon is rarely observes when global average is considered. In many outbreak scenarios there exist a non-source node with very high scope, thus playing crucial role for a specific outbreak origin. In this thesis we call such node a 'gate'. Gates are airports which have to be passed by traveling agents to reach a big fraction of the network and are therefore best targets for intervention deployment. A gate can take on this role for multiple out-

break scenarios and in fact most of the gates are hubs of considerable size. Exceptions arise when the outbreak is initiated at an isolated airport, having access to the WAN via a single connection to a bigger local airport. In this scenario the second airport becomes the major gate for the outbreak without having high flux itself.

In comparison to scope, confluence shows a similar trend, but there are multiple exceptions starring high scope and low confluence (e.g. London Heathrow, New York J.F. Kennedy and Los Angeles International) and vice versa (Hartsfield–Jackson Atlanta, Istanbul Atatürk and Chicago O'Hare; see figure 4.5). This is not surprising: on an effective distance tree a node, which has low confluence and a child with high scope, can have similar scope as a node with high confluence and hence high number of leaf node children. Nonetheless, it gives an interesting insight into the structure of the WAN, where a small node can become a gate by linking to a global hub. This case is rarely observed in the global average. An interesting example is Domodedovo Intl. airport (DME), Moscow, which displays second highest average confluence while having low average scope of approximately 1%. This airport is not base to any major airline but is a destination of several international and especially European airlines. The discrepancy can be observed on a much weaker scale in degree and betweenness as well. To our knowledge the difference can be explained by a high number of low traffic connections. While DME has a lot of children they account for low flux, giving it high confluence and low scope. Istanbul-Atatürk airport shows similar behaviour. Such discrepancies between scope and confluence might hint to specific network properties and are important for further investigation.

4.2 Application of the metrics on world aviation network

Scope of node n indicates the role of a node in a certain outbreak, its participation in the most probable transmission routes and the population fraction which can be reached through it. Higher scope means more infectious agents are likely to traverse the node, thus the node takes on a role of a gate into specific region. For this reason, high scope nodes are natural targets for intervention deployment like passenger entry or exit screening. Several papers have evaluated the effectiveness of exit screening with mixed results. While the potential of entry and exit screening to delay disease spread has been acknowledged in multiple publications [36, 27, 71, 13], current effectiveness of its implementations is controversial [89, 49, 59]. It has been emphasised that screening procedures are able to delay, but not prevent the spreading. It is most effective on early stages of the epidemic when numbers of infected are low [40]. Hence it is advisable to implement this countermeasures as early as possible in parallel with local containment at the outbreak location. This aligns well with the benefits of scope, which can be calculated very early knowing the origin only. Using WAN we investigate exemplary hub nodes and outbreak scenarios to highlight the application of the devised metrics.

In absence of an epidemic we want to characterise the airports to get a general understanding for the role of a node. As outlined above fully averaged scope is of limited value, but it is hardly possible to interpret the scope of a node for each outbreak location separately, since there are nearly 4000 possible outbreak locations. Thus we define a regional profile of an airport by averaging its scope across each of the 22 geographical regions. The regions are in accordance with the information provided by [58]. These profiles can be used to roughly judge the importance of an airport in case of regional outbreaks when no further information is available. Figure 4.6 shows the profiles of the biggest airport on each continent (**A**) and subsequent six biggest airports worldwide (**B**). Multiple depicted airports only play a role in case of an outbreak inside their own region. In this case most of the traffic leaving the region is directed through this node as it takes the role of a local hub. Many airports of this kind

are observed in the network in general and in this figure in particular. The most prominent example is Johannesburg O.R. Tambo (JNB). In case of JNB its importance in outbreak scenarios in the Southern African region is tremendous, with scope reaching about 71 % of the network. Further examples are São Paulo-Guarulhos and Sydney Kingsford Smith, the latter being important for two neighbouring regions. The scope of airports New York J.F.Kennedy (JFK) and Tokyo Narita (NRT) is dominated by only one region, but different than their own: Caribbean region in case of JFK and Micronesia in case of NRT. Both are small regions which use the respective airport as main transit hub. We can classify airports of this type as specialists, regardless if they are important for their own or other single or few regions. On the other hand there are generalist airports, which play an important role for many regions while never being the dominant transit airport for the whole network. Those airports shouldn't be neglected in any outbreak and can be safely considered important if no information whatsoever is available about the outbreak. Examples of such airports are London-Heathrow (LHR) and Dubai International (DXB). Scope of both airports varies across broader regions, e.g. low scope in case of outbreaks on American continent or higher scope in case of an outbreak in African and multiple Asian regions. Multiple airports lie in between both types: they behave as a gate for multiple regions, nevertheless the scope for remaining regions is non negligible. Examples of this mixed type are Beijing Capital (PEK) and Los Angeles Intl. (LAX) airports. PEK shows strong association with its own region being a central hub for China, but also uniformly increased scope for all other regions except Oceania and South- / South-East Asia. The latter combination is not surprising as airports from South-Eastern Asia often act as transit hubs for traffic from Oceania and vice versa. LAX has no single dominant region and multiple locations with scope close to zero. At the same time LAX gains high scope from many locations, mainly in Oceania and America. While clear distinction between those types is only possible in extreme cases, the information provided by the profile gives valuable insight into the role of an airport.

We already mentioned the most obvious types of profiles an airport can exhibit. To explore it further we clustered 100 airports with the highest total passenger flux according to their scope profile. The airports were clustered hierarchically based on mean correlation. Two approaches were employed. The result of the first is shown in 4.7 **B** and is based on profiles as demonstrated in figure 4.6, where the scope values are associated with geographical regions. To produce the clustering shown in 4.7 **A** the scope profile of each airport n war reordered so that $s_n(i) < s_n(i+1)$ where i is the index of the scope value. The airports were then clustered according to these modified profiles.

While the first approach associated airports by their region of influence, the later is an indication of the specialisation level discussed above. With the latter clustering approach airports are partitioned in two classes, reflecting the extremes mentioned above: generalists and specialists. Generalists are predominant among the 100 biggest airports, accounting for about two thirds of the airports. In a sense this is a disadvantage for airport administration, as generalists need further investigation at every outbreak. Nonetheless it is interesting that the airports are divided into two distinct clusters despite of the boundaries being smooth.

When airports are clustered using full profile information the resulting clusters can be easily tied to regions for which they have the highest importance. To assign a region to the cluster, scope profiles of all airports contributing to the cluster were investigated and the regions were ranked according to scope they grant to the airport. If a region was present in the top 4 of all but one or all airports of the cluster, the cluster was tied to this region. An exception to this rule is the mixed African cluster (purple): here the top 6 regions were considered, otherwise a classification was not possible. Detailed lists of region association can be found in Appendix 8.2. In several clusters the region of influence and geographical regions of airports overlap, e.g North-American cluster and Central Asian cluster. Here multiple

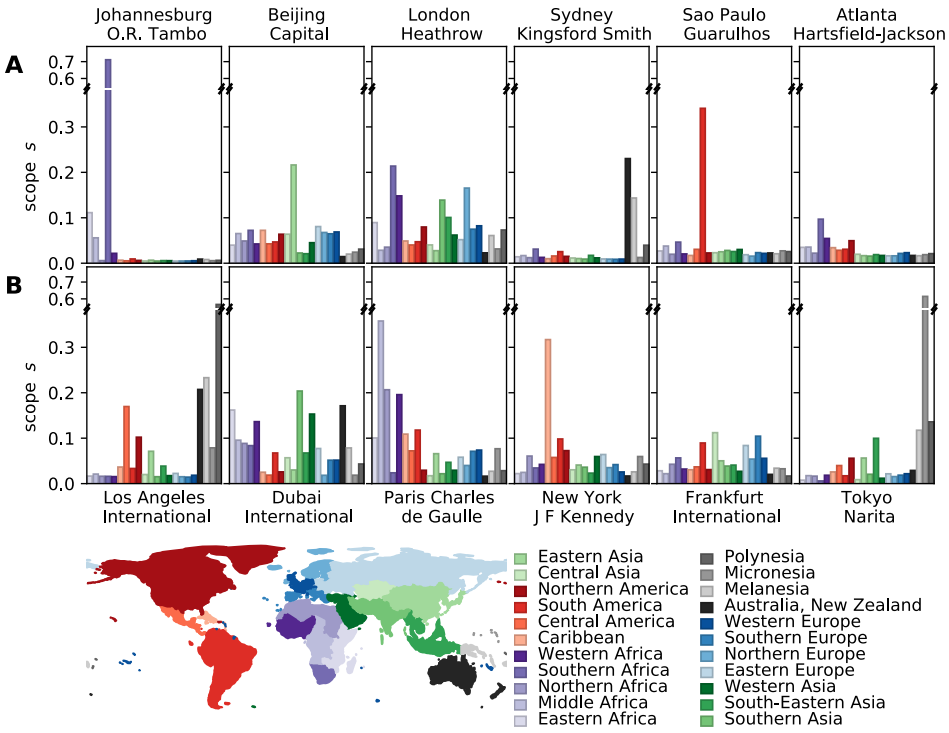

Figure 4.6: Scope profiles of big airports with respect to 22 regions of the air transportation network. The bars represent average scope of the airport with regard to different outbreak regions.(**A**) Biggest airports from each of the six continents. Some airports (i.e. Johannesburg O.R. Tambo and Sao Paolo Guarulhos) act as gates in case of an outbreak in the region they belong to, but play minor role when outbreaks happens in other regions. Other airports like Beijing Capital and London Heathrow play an important role for multiple regions. (**B**) Examples of hub airports displaying different profiles. Los Angeles International plays an important role for outbreaks in all 4 regions of oceania. Frankfurt International is a big hub in terms of passenger flux, but it has a relatively low profile for all regions and is not a gate for any of the regions.

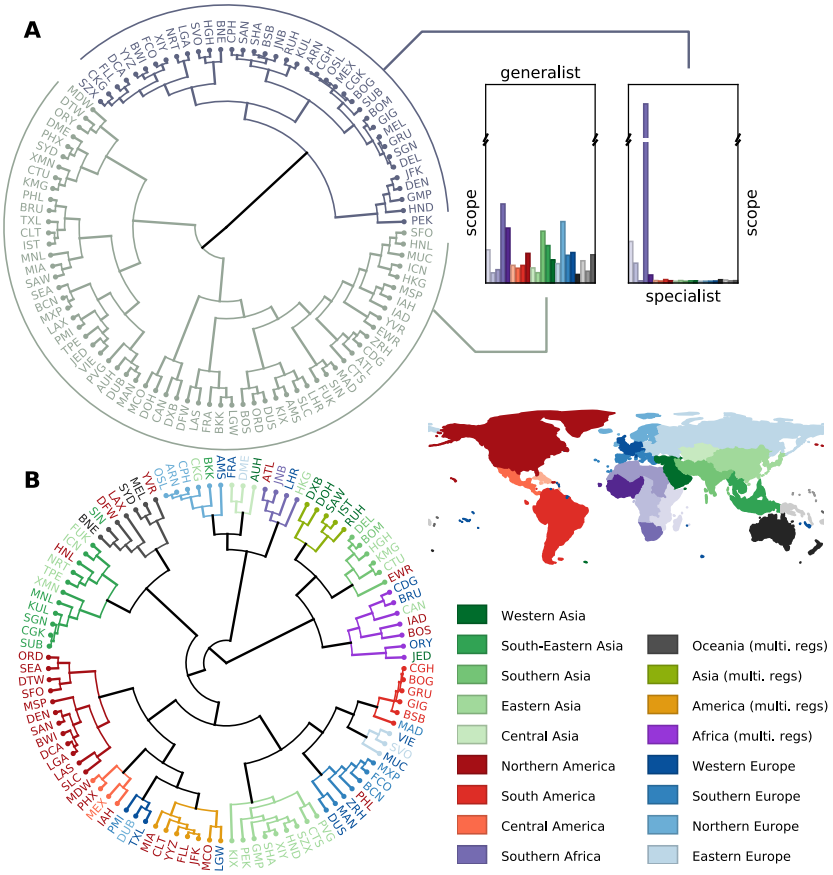

Figure 4.7: Hierarchical clustering of the top 100 airports according to their regional profiles. Final clusters were established using 'fclust' algorithm from scipy.cluster package. All branches inside one cluster share the same colour. Branch length shows the distance between the adjacent groups. (**A**) Clustering of the airports considering full profile information. Node labels are IATA letter codes of the airports and coloured according to the geographical region of the airport. Clusters are coloured according to the regions for which the most airports inside the cluster act as gates (see main text for more information). Note that the geographical location of the airport is not a determining factor for the assignment to a particular cluster. Best example is the African cluster (purple), which contains airports located in several geographic regions. Paris Charles de Gaule is inside this cluster and as we showed in previous figures is an important gate in case of an outbreak in Africa. London Heathrow, also shown to be an important gate from Africa, is assigned to a more specific South-African cluster (violet). (**B**) For this, clustering the regional information of the profiles were stripped and the scope values sorted. Airports were clustered according to the shape of the resulting profile. All nodes are partitioned into two clusters: generalists (dark grey) and gate (light grey). The gates play very important role for small number of regions. Examples of gate airports are Johannesburg O.R.Tambo, Sao Paolo Guarulhos and New York J.F.Kennedy. Generalist airports play a role in case of an outbreak in multiple regions but are never an exclusive gate into a region. Examples of generalist airports are London Heathrow, Paris Charles de Gaule, Beijing Capital and Dubai International.

smaller airports from the respective regions can be observed. Other clusters have influence in a broader region around their geographical location, e.g. the multi-regional Asian cluster starring airports from western Asia exclusively. In his case all airports combined in this cluster are big hubs and base airports of one or multiple airlines (with the exception of Sabiha Gökçen Intl. Airport, SAW). Many flight connections between Western and Asian regions are routed through airports in this cluster.

The most interesting clusters contain airports from a variety of geographical regions, which share the same region of influence. For example the multi regional African cluster with airports from Europe, USA and Asia. Some historical ties can be seen in this cluster: two French airports are present in it, accounting for the long history of French colonisation. Brussels presence in this cluster is also linked to the colonial ties it shares with Congo and Ruanda-Urundi. Other airports may or may not be attributed to history. Without a doubt all airports in this cluster accumulate a high fraction of traffic to and from Africa. This cluster highlights an important feature of infectious disease spreading. Outbreaks in certain geographic regions create very complex spreading patterns involving gates from distinct regions and continents. Such outbreaks inevitably require an international coordinated response to prevent a global outbreak.

Another example is the cluster associated with Oceania, in which two American and one Canadian airports are present. Two of these airports are located on the West Coast and are hubs used for transit flights from Australia and Oceania to multiple American and European destinations. The Oceania cluster also highlights strong association between Oceania and South-East Asia, represented by Changi Airport in Singapore. Another clusters highlighting historical relations are South-African cluster, which includes London-Heathrow, and Central Asia cluster, where the importance of Domodedovo airport, Moscow, may be attributed to the legacy of Soviet Union.

Scope was designed to be context sensitive and applied to different outbreak scenarios where detailed or limited information is available. Figure 4.8 **A** demonstrates this on real world outbreaks while figure 4.8 **B** outlines some hypothetical outbreak scenarios. The figure shows worldwide, top European and US American airports according to scope. We chose a subset of outbreak settings spanning different continents and sizes. In the worldwide ranking diverse airports are represented. In regional ranking, on the other hand, multiple airports occur in every scenario. For case import into European region those are Paris Charles de Gaule (CDG), London-Heathrow (LHR) and Amsterdam Schiphol (AMS). In every case either of the first two is being the top airport inside of the European region. The role of LHR and CDG is especially striking in case of the Ebola outbreak, where 50% of the network will most probably be reached through either of those airports. In this scenario the scope of LHR or CDG lies almost one order of magnitude above the scope of the airport ranking third ($s_{LGW} = 0.03$). Again we can observe how historical ties influenced the structure of the world air traffic. We can conclude that these three airports, LHR, GDG and AMS, are European top targets for countermeasure deployment if little information is available on the outbreak. They appear among the worldwide top 10 and are predominant gates in the European region. All of these airports are generalists with high capacity, thus their presence it the worldwide ranking of many scenario is not surprising.

A similar situation arises in the US ranking where New York J. F. Kennedy (JFK), Los Angeles Intl. (LAX) and Atlanta Heartsfield-Jackson (ATL) are present in every scenario with JFK ranking highest in many settings. JFK is present in the worldwide top 10 and is the biggest gate in case of a hypothetical outbreak in the Caribbean region. Similar to the situation with European gate airports, JFK, LAX and ATL can be valid targets for interventions when little information is available. Further, it is noteworthy that in two scenarios where either a European or American airport is taking on the role of the global gate the respective

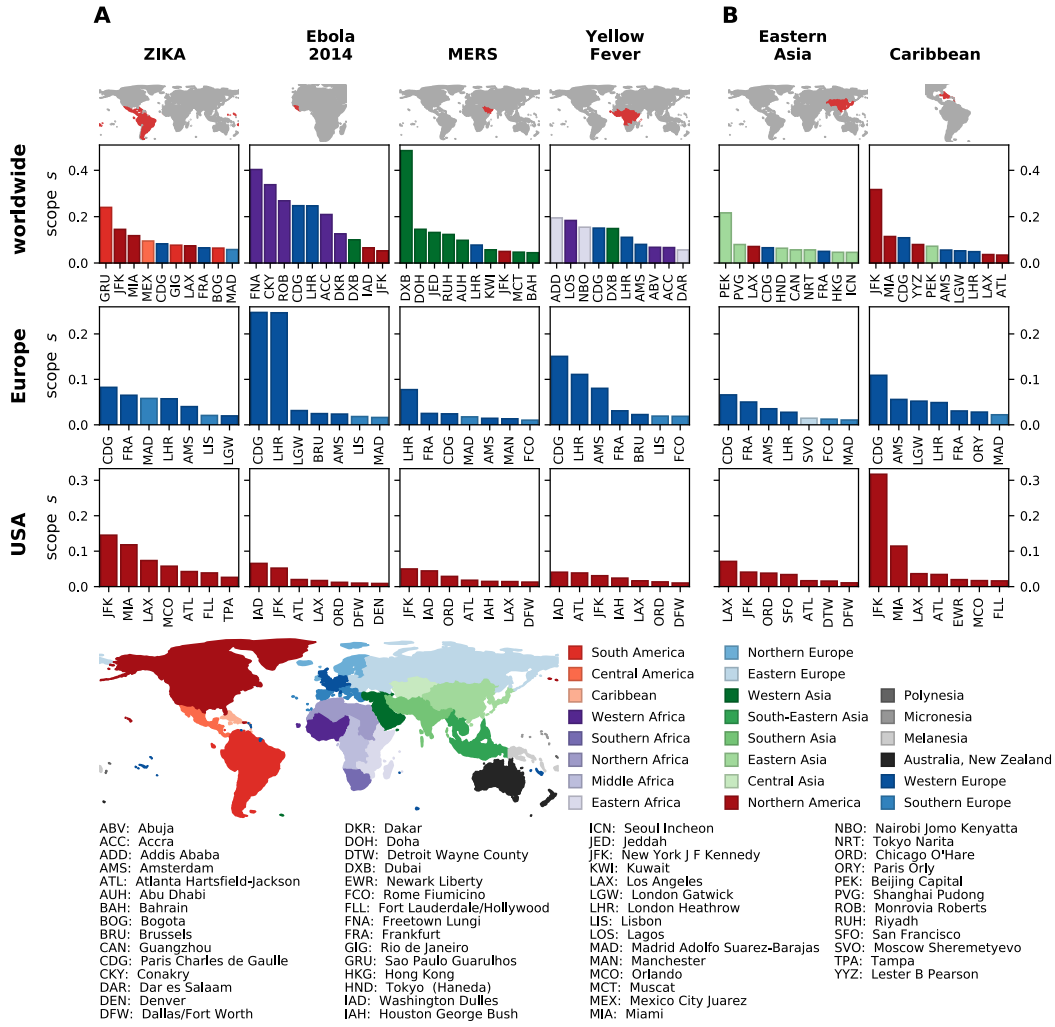

Figure 4.8: Application of scope on exemplary outbreak scenarios. Bars show the top airports with respect to scope in case of a real outbreak (**A**) and an outbreak in a geographic region (**B**). The top row shows top ten airports on the global scale, middle row - top 7 airports inside Europe, bottom row - top 7 airports inside the US. The outbreak region is shown on the map (red color). Bars are coloured according to the geographical location of the airport on the x-axis, labels are IATA letter codes of the airports. Different outbreak regions lead to distinct scope distributions with different airports being the main gate for the outbreak. At the same time the figure highlights that inside of a region there are important hubs which act as gates in a variety of scenarios, e.g. Paris Charles de Gaule and London Heathrow in Europe. In case of Ebola their role is especially prominent while there are no obvious gates among US airports. Reverse is true for an outbreak in the Caribbean region. Using proposed metrics countermeasures like passenger screening can be deployed in the gate airports, which lie on the most probable route from the outbreak location into entire regions.

other region exhibits rather low scope overall.

Worldwide ranking is showing a diverse picture for every scenario. Airports from multiple geographical regions, which are often different from the outbreak region, are starring in the ranking. Those airports act as gates into their own and other regions, distributing the disease worldwide. In case of ZIKA outbreak, the top gate is located inside the affected region, while the second largest gate is in the unaffected part of the USA. Multiple European airports are also represented in the worldwide top 10. In case of the MERS outbreak the top 5 gates lie inside the region affected by the disease. Those airports facilitate a big amount of traffic and act as worldwide traffic hubs. Hence it is not surprising that they take on the role of gates in this scenario. The dominant gate in this case is Dubai Intl. giving access to 49% of the networks population. This airport also enters the picture in context of Yellow Fever endemic region, side by side with multiple African airports. In the hypothetical scenario of an outbreak in Eastern Asia, PEK is the dominant gate, followed by multiple airports from the same region. While PEK accounts for only 20% of the scope, there is still a 10% gap in scope between it and the second ranking airport. In this scenario LAX and CDG are also present among the worldwide top 5.

Figure 4.8 demonstrates that there are some recurring patterns in the outbreaks and respective rankings. Some hubs always rank high when compared to other regional airports, often being distinct dominant gates. On the other hand when viewed globally, a dominant gate is difficult to predict using geographical information only. It often lies inside the affected region, but it is not guaranteed to, as can be seen in case of an outbreak in the Caribbean.

Until now we have concentrated on outbreak scenarios and characterised the role of nodes in context of the outbreak. Using scope it is possible to switch perspectives and focus on a node as a target rather than on a single outbreak origin. This is an important point when advising administration of an airport on matters of pandemic response with no regard to a specific pandemic. The aim of this application of scope is to raise awareness about locations relevant for the respective airport to facilitate a faster response in case of an outbreak. Again, different scales of averaging can be used for this approach. Often an outbreak is not limited to a narrow area and reporting is handled country-wise, hence it makes sense to evaluate threat at the countrywide scale. As can be observed in figure 4.9, all illustrated airports predominantly take on a role of major gates during outbreaks in the catchment area of small airports. Often those small airports do not share the same geographical region with the gate airport. An interesting example is Beijing Capital airport (PEK). The airport granting PEK the highest scope lies inside North Korea, highlighting the ties between these countries. This is also the only airport in this ranking which shares the region with PEK, while the rest belongs to Eastern Europe. Though this notation can be misleading: all these airports belong to the asian part of Russian Federation, which itself is attributed to East Europe according to WAN. Hence geographically listed airports are in close proximity to China. While being the airport with the highest average scope, PEK never acquires scope higher than 20%, except for an outbreak in Pyongyang. In fact, the decline in scope over all 3865 locations, when sorted, is slower when compared to other airports with high average scope like LHR or LAX. Hence we can say that PEK most often acts as a gate into Asian region, by that gaining high average scope. A similar picture arises when scope is evaluated country-wise: North Korea ranks first with a big margin to the rest of the ranking. The rest of the countries in the ranking are from different geographical regions including Africa, America, Europe and Asia. Russia is not present in the country-wide ranking, even though airports from Russia dominated the airport-wide ranking. The reason are the biggest Russian airports, which lie in the West of the country. They show weak association with China but influence the average heavily as they hold a big share of the overall population. Region-wise, East Europe appears as the region granting PEK a high scope. Eastern Asia does not appear in the ranking, but this

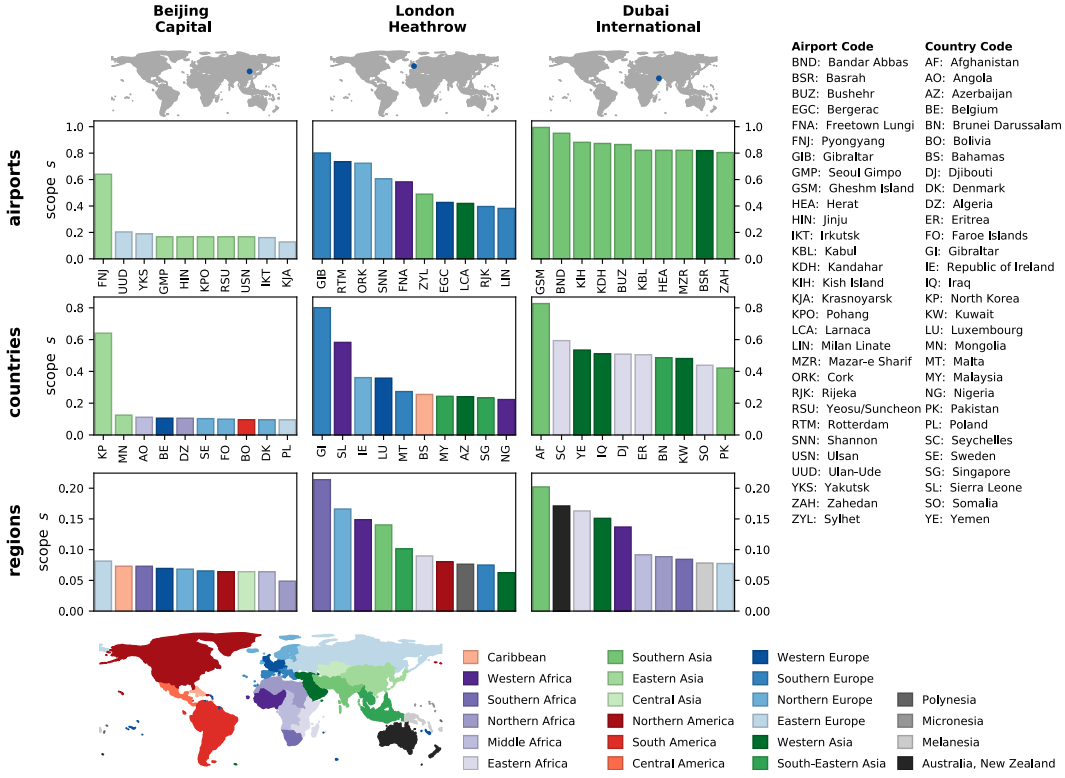

Figure 4.9: Relevant outbreak locations from the perspective of a hub airport. On the x axis are the top outbreak locations which lead to high scope values of the airport of interest (top label). Bars are coloured according to the geographical region of the outbreak location. Locations lying in the same country as the airport of interest were excluded from the ranking as we assume that close range disease import is more likely to happen via other routes. The top row shows the top locations in case of an outbreak confined to a catchment area of a single airport. The middle row shows the scenario in which entire countries are affected. The bottom row shows the region-wide outbreak and is identical to the profiles, previously described in fig. 4.6. Most of the important outbreak locations in the top row are small airports which route most flights via one bigger local airport. It is noteworthy that many outbreak locations lie in a different region than the airport of interest. Similar behaviour can be seen in case of a countrywide outbreak. Countries with low passenger fluxes route a high fraction of flights through a bigger hub, which becomes a gate to a considerable fraction of the network. This is illustrated by an outbreak in Gibraltar, in which case London Heathrow becomes a gate to 80% of the network. This change of perspective can help airport administration to better judge the threat posed by an outbreak in a specific region.

might be attributed to the filtering applied. For the plot, all locations from the same country were excluded as we assume that local transmission will more likely happen via different routes. In case of a country as big as China this assumption might be invalid. Furthermore, China is a substantial part of Eastern Asia, thus eliminating Chinese airports diminishes the overall impact of the region.

A very different picture arises when we consider Dubai Intl. airport (DXB). It is a base airport of Emirates airline and an international hub. It becomes a global gate with scope values approaching 100% for multiple outbreak locations in Southern Asia. Just as in case of PEK, airports granting DXB a high scope are small and isolated, with flights using DXB as main transit hub. Interestingly, no airport from Western Asia appears in the ranking with respect to DBX. This can be attributed to high concentration of hubs in Western Asia, hence the transit flights from inside West Africa are well distributed across multiple hubs. In a country-wise ranking West Asia is more represented, with Yemen awarding DXB about 50% scope. West Asia is also represented in the region-wide ranking, showing that while not being the only gate inside this region, DXB is well connected to West Asian airports. Multiple countries from Melanesia enter the ranking as well. This is also reflected in the region-wide plot where Melanesia ranks ninth. Compared to PEK or LHR this airport displays a very high overall scope in all top 10 rankings. All African regions rank high for DXB, especially Eastern Africa. South Asia is the highest ranking region for DXB, this is well in line with airport-wise ranking, which consists solely of South Asian sources, and the top ranking country, Afghanistan.

London-Heathrow shows a more diverse picture over all, displaying airports from multiple continents in the ranking. Freetown Lungi which has already been exemplified in 4.1 and as part of the Ebola outbreak in 4.8 ranks 6th. Consequently, Sierra Leone ranks second in the country-wise ranking with $s_{LHR} \approx 60\%$. Multiple European airports appear in the ranking granting LHR a scope of up to 80%. Despite of no airports or countries from Southern Africa being present in the ranking, this region shows a strong association with London-Heathrow in the ranking. Further regions from all continents are showing in the ranking, highlighting that LHR is an important international hub.

4.3 Discussion

As outlined above, scope and confluence provide new possibilities for informed decision making in the course of an outbreak. An airport starring high scope is a valuable target for passenger screening or comparable interventions. Hence, the resources can be concentrated on primary gates instead of a high number of nodes with possibly low impact. To ensure that our methodology can be easily applied by different interested parties the code used in this thesis will be available on GitHub as a Python / Julia package after the results have been published.

General classification of airports via profile and the clustering should give a general rule of thumb for the respective location, but we strongly encourage precise evaluation using the source in case of an outbreak. We have demonstrated how this method can be applied on a variety of real and hypothetical scenarios to determine gate airports. To calculate scope only little outbreak specific information is necessary, namely the outbreak origin. It is one of the easiest to establish parameters, as the export of cases via air traffic only becomes probable when the outbreak reaches a detectable size. Hence, we do not need information on case zero, but rather the location where the disease generates considerable number of secondary infections. This allows us to implement countermeasures in a timely fashion, minimising the probability of infected passengers to escape undetected before the screening or the like are

set in place. If deploying entity is a local administration of an airport, we demonstrated how the procedure can be adapted to classify possible outbreak origins, countries and regions important for the airport in question. Hence, a one-time counselling of the administration with respect to threats and important locations will be sufficient as long as the WAN is not undergoing major changes.

While global scope average looses some of its predictive power it still differs a lot from conventional centrality measures. It shows similar trends with established centrality measures, but for multiple nodes there is a considerable difference. The probabilistic argument behind the effective distance and thus the scope might explain the difference. By definition, scope not only considers the number of shortest paths crossing a node but also the size of the populations which lie downstream of the node. The latter is a more relevant information for public health, which seeks to protect the highest possible fraction of people. We need to point out that an assumption regarding the population size in the catchment area was made. While in generally correct, there might be a systematic difference in proportionality in developed vs. developing countries. If available, real census data should be preferred to estimate the population living in a catchment area of an airport. This data is subject to less systematic bias.

As has been demonstrated on Moscow Domodedovo (DME), Seatle-Takoma Intl. (SEA) and Istanbul-Atatürk airports, usage of both metrics, scope and confluence, can highlight topological specifics of the network. Especially DME and SEA host a high number of low traffic connections to low populated targets. Thus betweenness centrality and unweighted degree overestimate their roles. In contrary scope is able to differentiate between a global hub and airports hosting many low traffic connections.

In cases with a single known outbreak origin gradual averaging can be used to project how scope of nodes will be changing if a disease spreads locally. We need to consider that as soon as catchment areas of additional airports are affected new gates can arise. This allows administrations to be notified early and start countermeasure deployment and further preparations in advance. Nonetheless, we discourage averaging scope or confluence without a good reason to do so. As has been demonstrated in figure 4.1, both metrics might vary greatly across outbreak origins, even in close geographical proximity. Hence, averaging can obscure the real picture and lead to wrong decisions when no caution is applied.

Furthermore we want to emphasise that scope does not give any insight about the probability of an infected individual entering the node and interacting with the local population. When traveling by plane, transit connections are often a norm rather than exception. A passenger traversing a node in this manner will unlikely introduce the disease into the population of this area as the interaction time is low and a high fraction of the visitors of an airport do not belong to the local population. We are aware that transmission inside airport building and on a flight are possible and documented, but they are not a subject in this thesis [61, 74]. If we disregard the latter issue, despite close effective distance gate nodes might be at lower risk than nodes downstream from the hub, as they are more likely the terminal nodes of the journey. In the next chapter we will explain how to infer the probability that an agent will end his journey at a node from the topology of a network and derive a measure for the risk of disease importation.

Chapter 5

Quantitative assessment of import risks for pandemic onset situations

Following chapter introduces a methodology to calculate case importation risks in a particular outbreak scenario. Similar to scope and confluence, import risk relies on effective distance trees, and can be tailored to a specific outbreak scenario. Mathematical definition of import risk requires the knowledge of the probability that a traveller terminates his journey at a specific location, in short the *exit probability*. This probability often needs to be implied from the structure of the network. We provide a method for inference of the exit probability and demonstrate how it can be used to calculate the *import risk* from a specific outbreak origin. After describing the basic distribution of exit probability and import risk in simple, exemplary cases, we proceed by applying the method to hypothetical and real world outbreak scenarios. Furthermore, we show how import risk can be used from the perspective of a target location to establish outbreak sites which pose the highest threat.

When arguing about disease spreading, locally or globally, our thinking is often dominated by geographic distances. Remote outbreaks are perceived to pose little threat, regardless of the size of outbreak or specifics of outbreak location. Over many years this thinking was reinforced by observations, with most prominent and well documented example being Black Death. It swept over Europe in distinct wave-like front, with death toll estimates of about one third of European population of that time. Multiple other diseases followed a similar pattern [106, 101]. Following globalisation and advances in transportation wave-like patterns weakened and disappeared.

Severe acute respiratory syndrome (SARS) epidemic in 2002 was one of the first global outbreaks to demonstrate that geographic separation from outbreak source provides little security against importation of disease in the current age of fast global traffic. While the disease originated in China, the country with second most cases was Canada. A single infected from Guangdong province infected 16 local and international guests during his stay in a hotel in Hong Kong. Within few weeks after this event SARS was infecting people in 26 countries across 5 continents. Many countries with direct borders to China showed few cases with no sustained transmission, emphasising that reasoning in geographical means has little predictive power. SARS was a worldwide wake up call, forcing the WHO to issue a global alert [110]. The message that spreading is no longer confined by geographical borders was soon reinforced by a flu pandemic in 2009. Two years after SARS outbreak International Health Regulations (IHR) were signed by all member states of the WHO, granting it a framework

to access and coordinate international efforts in case of a global pandemic or an event with potential worldwide implications [109].

It was suggested by multiple sources that changes in global transportation are at the root of observed changes. Epidemic models, which account for worldwide air traffic, lie well in line with empirical data supporting this claim [55, 41, 25]. Recently it has been demonstrated that remapping geographical distance according to their probabilistic counterpart on the world aviation network (WAN), called effective distance, restores wave-like patterns. This methodology was successfully applied to several real world and simulated examples [17]. This result challenges conventional argumentation about importation risks: when wave-like patters arise on airports mapped according to effective distance, effective distance takes the same role the geographic distance played in past outbreaks. Hence properties of the transportation network determine the global spread and thus need to be considered with greater emphasis than plain geographic distance.

In this chapter we propose a method to estimate risk posed by an outbreak based solely on the topological features of the network and source of the outbreak. We will utilise the notion of effective distance from [17], which was explained in detail in chapter 3. This approach is tailored for transportation networks, which have several specific properties. First, most of the time only raw fluxes between two directly connected nodes are known and no information about the actual starting and end point of a trip is provided. Nonetheless this information is crucial. As has been demonstrated by SARS outbreak new epidemic hotspots are likely to be sparked at locations where infected are released in the native population. Knowing only raw fluxes we need to infer the probability that an agent will end his trip at a particular node. In this chapter we will describe one way of such inference. Using it we will further show how given an outbreak location import risk to every other node in the system can be calculated. Proposed metric shows the probability, that an infected who left the outbreak origin will end his journey at a specific node. As in the previous chapter we propose a way to account for limited information, spacial spreading and the severity of the outbreak. Finally we apply derived methods on the world aviation network (WAN) in real and hypothetical outbreak scenarios. Just as scope and fertility, import risk benefits from its context sensitive definition. We show how this metric can be used from a perspective of a non-outbreak node to rank possible sources according to the threat they pose.

5.1 Mathematical definition of import risk

As described above we require an agent to exit at a node to be able to spark an epidemic. Hence we are not only interested which link will be taken by the walker next, but whether the walker will continue at all. We are given a transportation network with N nodes connected by weighted links $F_{mn} \geq 0$. Usually raw fluxes are the only information available about passenger movement. Nonetheless we want to know the probability $p(n_L, n_0)$ that an agent starting at node n_0 travels to node n_L and exits there. We call this probability import risk . There are multiple paths between two nodes, infinite if loops and multiple transitions over the same node are allowed. In real air traffic networks passengers are unlikely to exhibit either behaviours, so we require the paths to be self avoiding. Import risk has two parts to it: the probability that an agent travels along specified links and the probability that he exits at n_L and not before. We call this probability exit probability q_{n_L}. For the moment we assume to know the latter and proceed with the import risk definition.

Now consider a single path $\omega = \{n_0, n_1 ... n_L\}$ with a walker starting at n_0, traversing all the nodes of ω and exiting in n_L. In previous chapter we derived the probability that an agent at node n_0 makes a step to its direct neighbour n_1, $P_{n_1 n_0}$ 3.1. Using it we can calculate

the probability of a walker to traverse ω and exit thereafter:

$$\begin{aligned} p(\omega) &= (1-q_{n_0})P_{n_1 n_0}(1-q_{n_1})P_{n_2 n_1}...(1-q_{n_k})P_{n_k n_{k-1}}...(1-q_{n_{L-1}})P_{n_L n_{L-1}}q_L \\ &= q_{n_L}\prod_{k=1}^{L} P_{n_k n_{k-1}}(1-q_{n_{k-1}}) \\ &= q_{n_L}\prod_{k=1}^{L} S_{n_k n_{k-1}} \end{aligned}$$

where $S_{n_k n_{k-1}} = P_{n_k n_{k-1}}(1-q_{k-1})$. While we could argue that import risk is dominated by the probability of importing disease over the most probable path or path with the shortest effective distance there is a myriad of less probable path which contribute to import risk. Some of those paths can be only slightly lower in probability than the shortest path and sum up to very high import probability in addition to shortest path. Hence we should consider a bigger subset of paths. We can extend the formula to calculate the probability of a walker starting at node n_0 and exiting at n_L by means of all paths of length t

$$p(n_L, t|n_0) = q_{n_L}\sum_{\omega\in\Omega_t}\prod_{k=1}^{L} S_{n_k n_{k-1}}$$

where $\Omega_t = \Omega(n_L, n_0, t)$ is a set containing all possible paths from n_0 to n_L within t steps. Note that we are talking about length in terms of steps, not in terms of probability. Consideration of path length can become relevant if the timescale of disease and temporal scale of steps are of same magnitude. We will not examine this special case in the work at hand as for most diseases and the air traffic timescales are orders of magnitude apart. Assuming infected can not recover during the journey irrespective of its length we are interested in $p(n_L, n_0) = p_\infty(n_L, n_0)$, the probability that an agent leaving node n_0 arrives at node n_L regardless the path length.

$$\begin{aligned} p_\infty(n_L, n_0) &= \sum_{t=1}^{\infty} p(n_L, t|n_0) \\ &= q_{n_L}\sum_{t=1}^{\infty}\sum_{\omega\in\Omega_t}\prod_{k=1}^{L} S_{n_k n_{k-1}} \\ &= q_{n_L}\sum_{t=1}^{\infty}\left(S^t\right)_{n_L n_0} \end{aligned} \tag{5.1}$$

From above definition follows $\sum_m p_\infty(m, n_0) = 1$. The notion S^t describes the probability that an agent will travel from its starting node to a target node in means of t steps without exiting on his way. The derivation of this matrix notation is explained in more detail in [70]. Calculating import risk in this fashion can become computationally challenging on big networks. We can devise methods to limit the number of paths to a smaller subset. With increasing length import probability over a path is strictly declining, hence all longer paths contribute less to import risk. In fact $p(m, t|n_0) \to 0$ when $t \to \infty$ even though the number of possible path grows as the number of steps increases. We can thus consider paths only up to a maximal length, which significantly decreases computation time by removing a high

amount of paths with minor loss of import probability. In mathematical terms

$$p_{max}(n_L, n_0) = \sum_{t=1}^{t_{max}} (S^t)_{n_L n_0}$$

On some highly heterogeneous networks paths of same length in terms of steps can have very different probabilities. In this case we can introduce a probability threshold, below which a path is not considered. Similarly effective distance can be used. This is mathematically less straight forward and can not be expressed in matrix notation easily, but it can be implemented in a programming language of choice using modified Dijkstra algorithm. This approach was used in a later chapter 6. For more details on the implementation see chapter 2.

Until now we assumed to know the exit probability. WAN data does not contain any information about the probability of an agent to end its journey at a particular node. In most if not all transportation networks the number of passengers travelling between two remote, not directly connected nodes, is unknown. Thus we need to estimate exit probability. The easiest way is to assume equal exit probability q across all nodes. Using the relation $\sum_m (P^t)_{mn_0} = 1$ and the prior assumption about exit probability to calculate the probability of a walker to have not ended his journey after t steps

$$p(t|n_0) = q(1-q)^t$$

Using this formula we can determine how fast import probability over all nodes approaches one. Using this as a sanity check we can calculate $\sum_t^\infty p(t|n_0) = q * 1/q = 1$. This formula can be used to pick a sensible threshold to reduce computational time. Nonetheless, uniform exit probability is a strong assumption. It is likely that exit probability at a transit hub is considerably lower compared to a small remote airport which is rarely used for transit. Hence, we need to account for the role a node plays in a current outbreak. In previous chapter we outlined that scope s_n is a measure of how much population can be reached through a node acting as a transit hub. Node n with $s_n = c_n$ is a terminal node, hence we expect everyone who arrives to exit here. Nodes with $s_n \gg c_n$ are used for transit and agents are less likely to end their journey here. As previously outlined the role of a node and thus its scope is highly dependent on the outbreak origin. And so is the exit probability. Aside from node's scope, node's population also has an impact on exit probability. We assume that $q_n \sim c_n$, meaning that high population nodes are attractors with higher exit probability while transit nodes with low population are more often solely traversed.

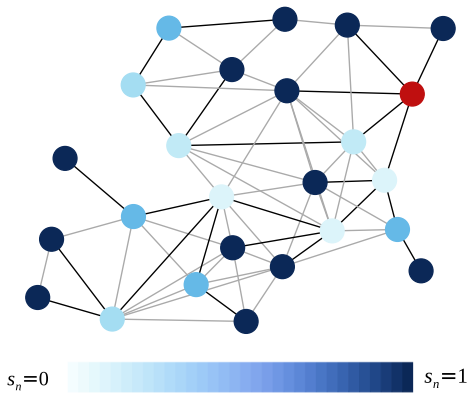

Figure 5.1: The distribution of exit probability across a Delaunay network. Red circle marks the outbreak origin. The colour of the remaining nodes indicates the exit probability at the respective location with darker shades representing higher exit probability. Black links are part of the effective distance tree. Along a branch of the effective tree the exit probability increases as the distance from the origin increases. Leaf nodes show highest exit probability, $s_n = 1$.

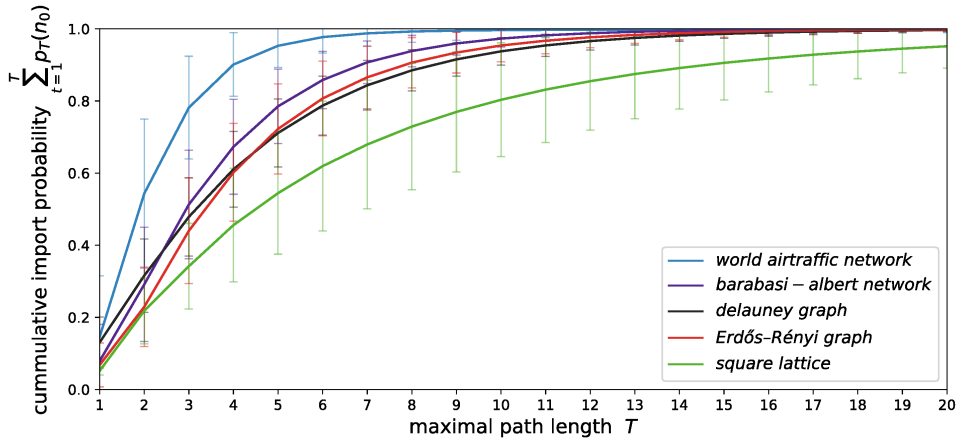

Figure 5.2: Total import risk over all nodes depending on the maximal path length. Solid line shows the mean of cumulative import probability, shaded region - one standard deviation. For synthetic networks number of nodes $N = 100$, flux over each edge was assumed to be one. Captured import probability approaches a limit value of 1 for all network types. In highly inhomogeneous networks cumulative import risk saturates faster, while on regular networks, e.g. lattice, long paths need to be considered to account for high proportion of import risk. In the rest of this chapter $\tau = 14$ was used.

This leads to definition of exit probability as

$$q_n(n_0) = \frac{c_n}{s_n(T_{n_0})} = \frac{c_n}{c_n + \sum_{k \in \Theta} c_k}$$

where $\Theta = \Theta(n, n_0)$ is the set of children, grandchildren etc. of n on the effective distance tree T_{n_0}. We can assume same proportionality as in previous chapter $c_n \sim f_n$ and hence $q_n = f_n / s_n$. In this thesis we will use this scope dependent definition of exit probability. Note that the estimate of exit probability is the only part of import risk which relies on effective distance trees. The above definition is one of various possibilities and a heuristic estimate. Therefore it is meant to approximate exit probabilities when real values are unknown. We encourage to use real world data on start and destination if available. Regardless of what exit probability estimate is used import risk depends strongly on the outbreak origin. It follows the proposed logic of a context sensitive metric and can be applied on early stages of an epidemic to guide decision making.

While mathematically tractable ∞ is not a feasible value for computations. To decide on maximal path length to be considered τ we run simulations on multiple network types as shown in Fig. 5.2. Synthetic networks were generated to have the same flux over every edge, absolute value was irrelevant as all link probabilities are computed as fractions of fluxes. In highly inhomogeneous networks cumulative import probability approaches 1 faster. There are multiple reasons for this. First, in networks with small world features all nodes can be reached in means of very few steps. Hence all shortest paths are captured even with low τ. Second, after reaching a hub n with many neighbours the probability to proceed to any special neighbour P_{kn} is low for every k (though still $\sum_k P_{kn} = 1$). As has been demonstrated that by picking a random link you are more likely to reach a hub than a small node [70]. Hence following this logic the path probability decays very fast in an inhomogeneous network. Lattice as the most regular network has cumulative import probability well below 1 even when

paths of length 20 are considered. In this chapter we fix $\tau = 14$. It gives reasonably good results while keeping computation time low.

As has been demonstrated with scope import risk can be averaged in the same fashion. If multiple locations are affected by the disease, constituting a set Y, a weighted import risk can be calculated as

$$p_\infty(n_L, Y) = \sum_{i \in Y} \alpha_i p_\infty(n_L, i) \tag{5.2}$$

where $\alpha_i = c_i / \sum_{k \in Y} c_k$. For more detailed explanation see page 38. The same method can be applied on exit probability, but we in general discourage to do so. Exit probability varies more than any of the proposed metrics, hence there is a considerable risk that the average value will greatly misrepresent the matter at hand. It can be interesting to calculate cumulative risk to a wider geographical or political region. Assume Λ is a set of airports which belong to a single country, thus the import risk into the country is

$$p_\infty(\Lambda, n_0) = \sum_{k \in \Lambda} p_\infty(n_k, n_0)$$

This aggregation can be extended to broader geographical regions and continents. Both, averaging and aggregation, can be combined.

As shown in Fig. 5.3 there are general trends to both, exit probability and import risk with respect to the distance from outbreak origin. The figure shows an effective distance tree of a network with nodes coloured according to their exit probability and import risk. Exit probability grows with increasing distance from outbreak origin, for terminal leaves $q_n = 1$. Despite the general trend, nodes lying on the same radial distance often have different exit probability. This behaviour can be attributed to highly different scope when nodes lie on distinct branches of the effective distance tree. Further exit probability is not guaranteed to increase or remain equal along one branch, even though scope is monotonically decreasing. For detailed derivation of special cases see Appendix 8.3. In Fig. 5.3 import probability shows inverse trend, decreasing with growing distance to outbreak origin. There are two components contributing to import risk: q_n, which generally increases with distance on the tree, and $\prod_k S_{n_k n_{k-1}}$ which decreases with growing distance. Just as in case of exit probability import risk can exhibit more complex behaviours than shown in Fig. 5.3. We will demonstrate using WAN that risk of import for gate nodes, which are usually close to the outbreak origin, is often lower as for direct children of the gate.

Import risk and exit probability can be compared to other centrality measures when averaged over all possible outbreak origins in a network. Fig. 5.4 shows how import risk relates to other centrality measures. As has been outlined above exit probability and import risk show opposite trends and low conformity despite of $p_\infty(n) \sim q_n$. The reason for opposing trend can be explained by the relationship between import risk and scope. As can be seen in the figure nodes with high average import risk show high average scope spanning several orders of magnitude. The reason is that both metrics depend highly on the effective distance from the origin. Exit probability relates to scope as $q_n \sim 1/s_n$ and hence having a trend opposite to import risk. Betweenness centrality and import risk show little consistency, hinting that betweenness centrality is a bad predictor for import probability. Within 100 biggest nodes in terms of flux (highlighted in the figure) little can be said about the dependency between all metrics. Those nodes show the biggest variance in importance across different trees. Hence averaged values are little conclusive.

Some more explanation about the exact meaning of import risk p_∞ is needed. The probability $p_\infty(n_0, n_L)$ is the probability of an agent exiting n_0 to arrive and exit at n_L. Note that no statements are made about the likelihood or the number of agents starting at n_0 in the first

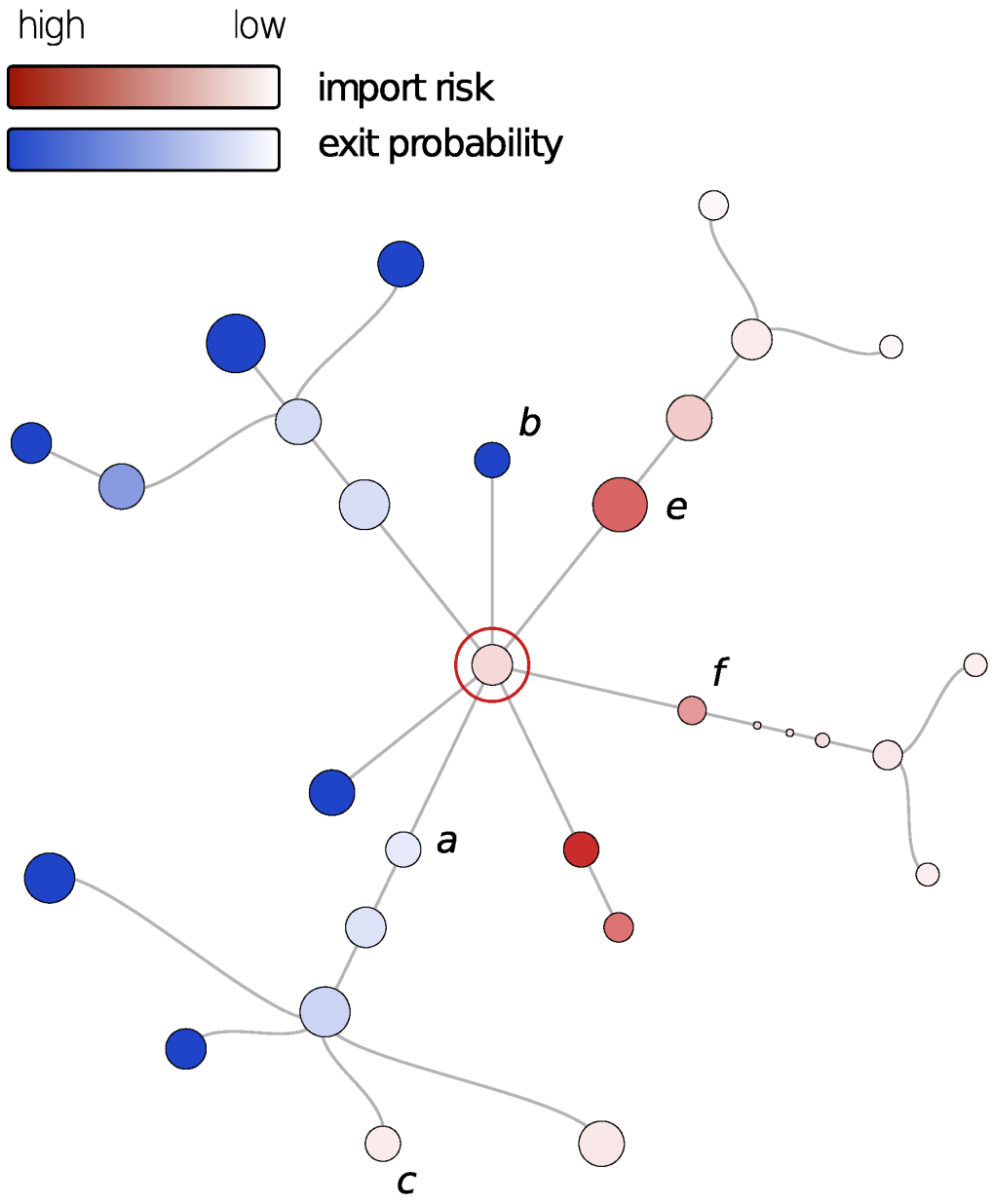

Figure 5.3: Example of the distribution of import risk p_∞ (red) and exit probability q (blue) in a network. Depicted shortest path tree was derived from a random geometric graph, node enclosed by the red circle is the outbreak origin n_0. The distance of a node from the centre is proportional to the probability of an agent leaving the outbreak origin and arriving at the node (but not necessarily exiting). The size of the node scales with its population c_n. By definition $q_{n_0} = 0$. For all other nodes n $q_n \sim c_n$ and $q_n \sim s_n$, for terminal nodes n_l $q_{n_l} = 1$. Some dependencies can be seen on nodes **e** and **f**: $q_e > q_f$ while $s_e > s_f$. The difference in exit probability can be explained by $c_e \gg c_f$. For import risk $p_\infty(n|n_0) \backsim q_n$ and $p_\infty(n|n_0) \backsim D_{eff}(n|n_0)$, which makes it non-trivial as $q_n \backsim 1/D_{eff}$. Nodes **a**, **b** and **c**: $c_a \approx c_b \approx c_c$, $q_b = q_c = 1$, $D_{eff}(a|n_0) = D_{eff}(b|n_0)$, $p_\infty(b|n_0) \gg p_\infty(c|n_0)$ and $p_\infty(b|n_0) \gg p_\infty(a|n_0)$. Having short effective distance and high exit probability b is at highest import risk. While $a$is close to the infected its low exit probability reduces the risk posed by the infected.

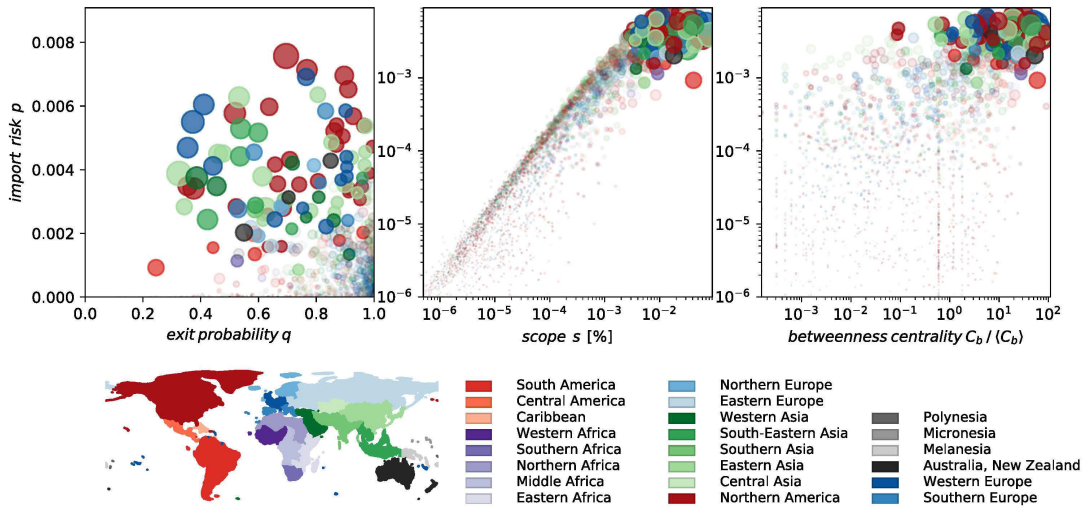

Figure 5.4: Relationship between import risk p_∞, exit probability q, scope s and betweenness centrality C_b. Nodes are coloured according to the geographic region they belong to, 100 biggest nodes flux-wise are highlighted. Size of the node is proportional to the passenger flux through the node. None of displayed metrics is a good predictor for import risk.

place. This is an important distinction to make when real threat of import has to be accessed. Let us consider a disease which has no latent period and renders the patient immobile or kills him in short time. The probability of a person infected with such disease to board a plane will be close to zero, hence the calculated import risk will be overestimating the risk. A different disease with long latent period, mild symptoms or imposing little restrictions on the infected will be likely exported from the area more often as the passengers will perceive to be healthy. In this scenario import risk will likely underestimate the threat.

5.2 Application on world aviation network

As outlined in the introduction global spread of diseases is dominated by the air traffic. Hence we apply the metric on the world aviation network (WAN) to demonstrate how proposed methodology can be used in a real world pandemic scenario. Most of the times outbreaks overarch multiple cities and spread locally so that import risk from one airport is not sufficient to characterise the threat. In this subsection we demonstrate how averaging can be used in this case and to compensate for lacking information. We will demonstrate the procedure on well documented outbreaks from recent years and on some hypothetical scenarios. The context dependent feature of import risk has further advantages: a switch in perspective is possible if the interest lies on import target rather than the outbreak location. This can be the case for administration of a certain airport who wants to know what outbreak locations would pose a threat to their area of responsibility.

When no particular outbreak is threatening the network general classification of airports according to import risk is possible. Average import risk over all possible outbreak and target locations in the WAN is $\bar{p} = 0.00026$. Due to the heterogeneity of the network and the context sensitivity of import risk, this value is not representative for any particular scenario. Fig. 5.5 **A** shows top 20 highest ranking airports in terms of average import risk from all other nodes is calculated. Import risk of the most threatened airport is more than 20 times higher than

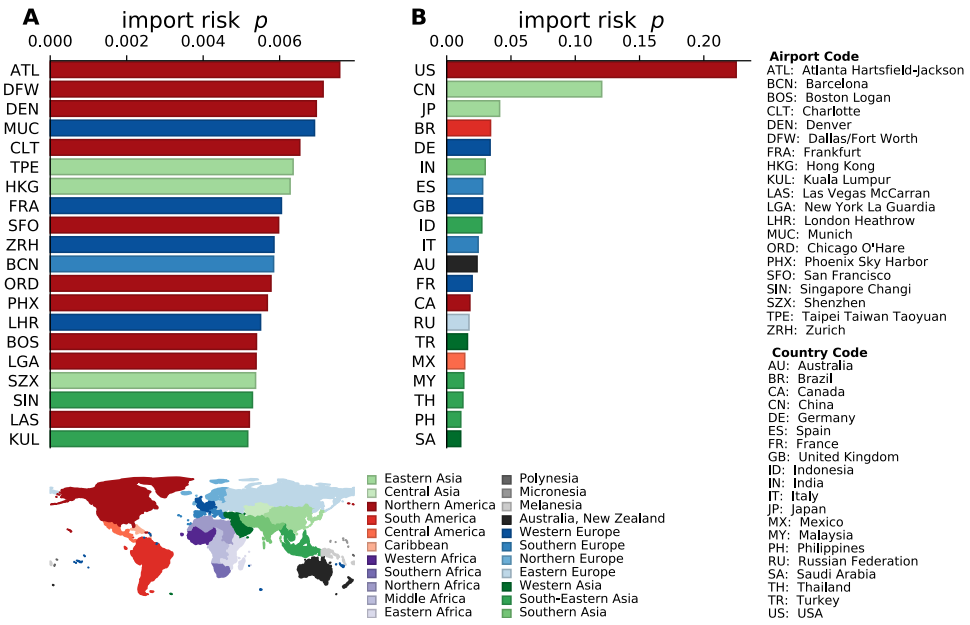

Figure 5.5: Top 20 airports (**A**) and countries (**B**) according to global import risk. Global import risk is the average import risk over all possible outbreak location weighted with the size of the outbreak location. The airports are labeled according to the IATA 3 letter code, country labels are ISO two letter codes. Bars are coloured according to the geographical region of the location. Only 4 out of top 20 airports in import risk ranking are present in top 12 of the scope ranking in Fig. 4.4. Airports in question are London-Heathrow, Frankfurt Intl., Singapore Changi and Chicago O'Hare. This demonstrates that the two metrics, p_∞ and s, have different meaning: scope indicating important transit nodes and import risk showing location in danger of disease importation. In country ranking the population of the country has an impact on the ranking, but it is not a determining factor. Multiple countries with low population are present in the top 20. Examples are Malaysia (ranked 42 according to population), South Africa (population rank 25) and Spain (population rank 30) [10]

average. Nevertheless absolute values of import risk for a single airport are low, all shown airports have import risk values of a similar scale. Most at risk airports are medium hubs from North America, Europe or Asia. There is little overlap with top 12 scoring airports according to scope (4 out of 12; Fig. 4.4) and betweenness centrality (3 out of 15; Fig. 4.2). Moderate overlap can be observed with flux (7 out of 15; Fig. 4.2) and eigenvalue centrality (8 out of 15; Fig. 4.2). We propose that primary gates on the effective distance tree are at lower import risk than their direct children. Primary gates show much higher scope than their offsprings while population size can remain comparable. As discussed in the previous chapter they are used as transit location, which leads to lower exit probability and hence decreased import risk. While primary gates are ideal locations for enter or exit screening, the secondary neighbours are at higher risk of case importation. Despite high exit probability small local airports don't enter top 20 as the probability of an agent travelling to the terminal leaf without exiting on the way is minor. As has been described above import risk is subject of two opposing effects: probability of an agent getting to the node, which decreases with the number of links traversed, and exit probability at the node, which increases the farther from origin the agent has travelled. Hence maximum import probability lies somewhere in between primary gates and terminal leafs.

Country-wise import risk is shown in Fig. 5.5 **B**. Difference in import risk between countries is more significant compared to the difference between airports, which can be in part attributed to difference in country sizes and the number of airports contributing to it. Nevertheless the size of a country is not the only factor, as multiple small countries appear within top 20, e.g. Malaysia, Saudi Arabia and Spain. All of the listed countries are well developed in means of economy and access to air transportation network. All of them host global and regional hubs hence accumulating disproportional amount of air traffic. Same cases can conflict with the assumption that traffic flow is proportional to the inhabitants inside the catchment are. In that case $c_n < f_n$, hence the exit probability for this node is being overestimated as is its contribution to scope of its parents. Reverse case can also happen. An example of it is Africa: while in part densely populated, amount of airports on African continent is comparably is low. In some regions of Africa population over all is less mobile in terms of air travel as has been discussed in Section 3.1. We acknowledge this limitation but don't think that it renders the overall method invalid. When crucial this problem can be circumvented by use of real census data.

In countrywide ranking more geographical regions are present. Australia and Russia are present in top 20 threatened countries on place 11 and 14 respectively. While no single airport from either shows in the airport ranking both countries extend over a wide geographical area and accumulate many regional and international airports (153 in Russia and 136 in Australia). Hence for both countries the risk of an importation from any specific airport might be negligible, but accumulated over all possible importation targets the threat is substantial.

More refined picture arises when we differentiate the risk posed by different regions instead of taking a worldwide average. This way we can derive profiles for airports and countries similar to those used in Chapter 4. Fig. 5.6 shows profiles for the biggest airport (**A**) and respective countries (**B**) for each of 6 continents. None of depicted locations is dominated by one region in terms of import risk, which is in sharp contrast with scope (Fig. 4.6). In a sense all airports are generalists when it comes to import probability. Most countries and airports are threatened significantly by outbreaks on their continent. While often enough biggest hub in a region acts as a major gate in case of an outbreak it is also guaranteed to be at close effective distance to all airports in the region. This leads to higher average import risk across own region. For a country the effect is even more profound as smaller local airports are very likely to be direct children of a major local hub, hence having a high import risk.

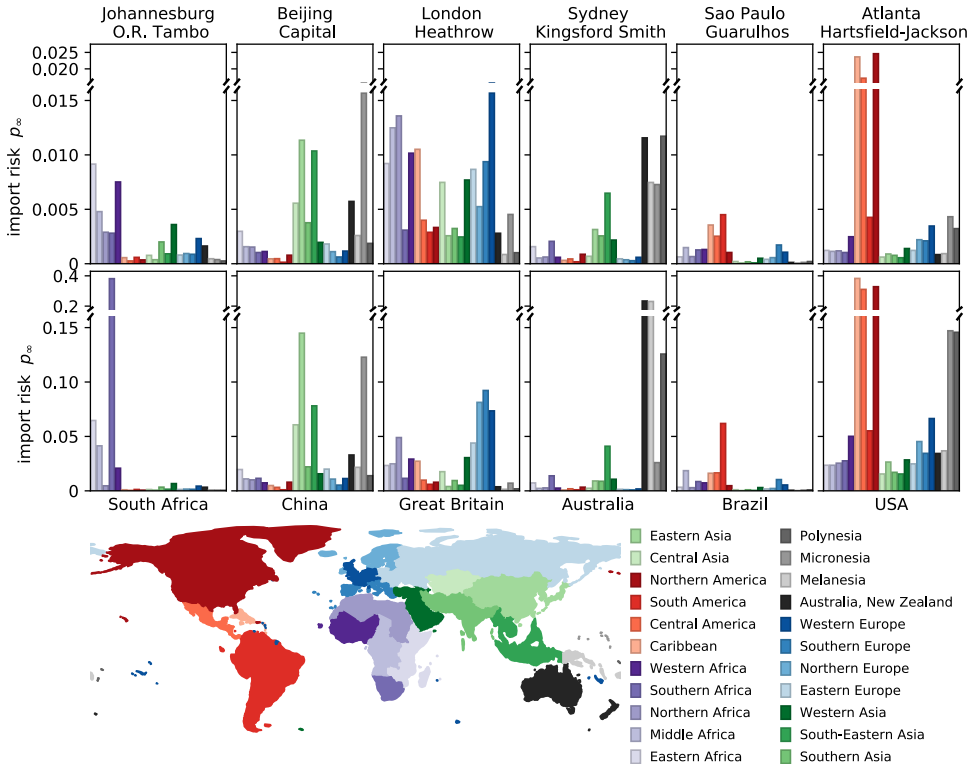

Figure 5.6: Import risk profiles of the biggest airports (top row) and respective countries (bottom row) for each continent. Bars represent regions and are coloured accordingly. Location itself was not included in the region sum. Different profile types can be observed: airports which have an overall high importation risk (e.g. London-Heathrow) or airports with only few locations posing high risk (e.g. Atlanta Hartsfield- Jackson). The country profile is heavily influenced by its biggest airport. An exception is South-Africa, where the highest peak, region of southern Africa, originates from smaller local airports. In the United States highest import risk can be attributed to its own continent. Airports on the west coast contribute to the considerable import risk from Oceania, this is reflected in the countrywide risk profile.

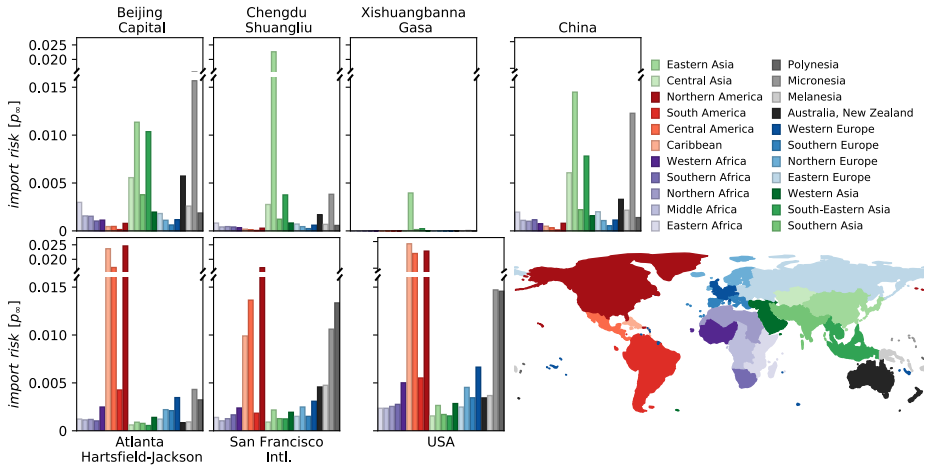

Figure 5.7: Country profiles of China and USA with exemplary airport profiles. Bars are coloured according to geographical regions, bar height indicates the import risk from respective region. Airport profiles inside one country display similar risk across multiple regions, e.g. high threat is posed to all Chinese airports by eastern and south-eastern Asian regions. This is the case for most of the 188 Chinese airports. Hence most airport profiles inside China match profile of the biggest airport and result in a similar country profile. In United States airport profiles overlap for American regions but can differ in import risk from Oceania. This is reflected in the countrywide profile.

Profiles of most biggest airports align well with profile of the country it belongs to. Prominent example of this is China, where country profile is strongly correlated to profile of Beijing Capital airport (PEK). It is even more striking as there are total of 188 airports in China. All Chinese airports are at high import risk from East Asian region, multiple big Chinese airports have a profile highly similar to that of Beijing Capital. These effects combined result in a country profile highly resemblant of its main airport as can be seen in figure 5.7. For Atlanta Hartsfield-Jackson airport (ATL) highest threat is posed by multiple regions of American continent, there is only minor association with Oceania. While ATL is the largest airport in WAN and America according to passenger flux there are multiple major hubs inside the US with a distinct profiles. An example of an alternative hub is San Francisco Intl. airport (SFO) which is at high risk of import from Micro- and Melanesia. ATL and SFO also represent typical airport profiles of the east and west coast respectively. It is not surprising that the country profile of USA is a combination of both, exposed to high risk from both American continents and Oceania (see figure 5.7). In big countries like the US a dominant airport is difficukt to determine, especially because of big geographical stretch of the country which is also reflected in connections of the WAN.

Another interesting relationship can be observed when the profile of Johannesburg O.R. Tambo (JNB) airport is compared to its scope profile in figure 4.6. The scope profile is dominated by a single region, South Africa, while import risk this region poses is significantly lower than the threat from other African locations. This underlines the conclusion that primary gates are not the most endangered locations. When JNB is the sole gate, which is the case for Southern Africa, probability that a passenger will exit in JNB is minor. In terms of flux and thus the population JNB ranks 82. When its scope is close to 1 its population is not sufficient to increase exit probability to high value. For other regions of Africa JNB is an

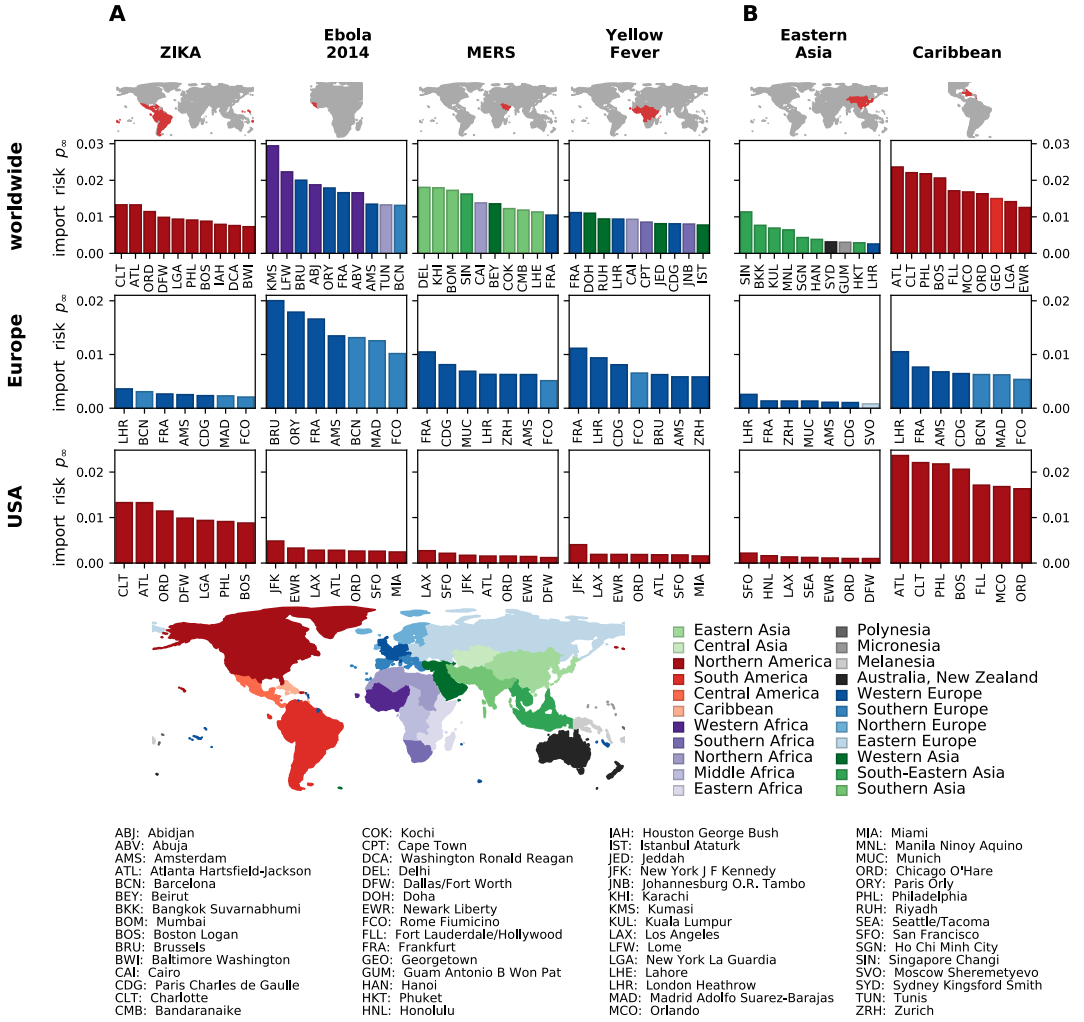

Figure 5.8: Application of import risk on real (**A**) and hypothetical (**B**) outbreak scenarios. Red region on the mini map highlights the outbreak region. Airports located inside the affected region were excluded from the bar chart. The airports are labeled according to the IATA 3 letter code, country labels are ISO two letter codes. Top row of the bar charts depicts most endangered locations worldwide, middle row - locations in Europe, bottom row - locations in United States. Top ten locations in each category are depicted. The bars are coloured according to the region to which the airport belongs. Some scenarios, e.g. Ebola outbreak or Yellow Fever, pose much higher threat on the European region than on United States. Or vice versa in case of Zika. In multiple scenarios it can be seen that the endangered locations lie in a different geographic region than the outbreak itself. In case of Ebola multiple European airports are in the worldwide top 10. Top three worldwide locations at risk with respect to MERS lie in Southern Asia, two of them in India.

airport with many scheduled flight but is transit for only minor airports inside south Africa. Hence it is a secondary gate with comparably high population, which results in a high exit probability and due to its position on WAN also short effective distance.

When more information about the outbreak is provided import risk can be calculated

to reflect the situation at hand. We incorporate the information about geographic spread through a weighted average over all airports inside the affected area. This approach does not consider that some regions can be affected more than others, but precise, quantitative information about the severity of the outbreak is rarely available and up-to-date. Nonetheless our metric can be easily adapted to incorporate this information by defining an altered average. Note that the import risk is not to be seen as an absolute value of importation probability in case of pandemic. Import risk evaluates the probability of a single agent, which left the outbreak origin to finish his journey at the respective airport. In a real world outbreak it is difficult to evaluate how many infected citizens will leave affected region as it depends on the mobility of the population and the disease itself. In countries where air traffic is underdeveloped or in cases when disease puts a great strain on infected right after onset probability of an infected to travel is greatly diminished. Hence we advise to see import risk as a metric to compare danger posed to different locations not as an absolute probability as long as the number of infected leaving the outbreak site are not accounted for.

As outlined above broader geographic regions affected by disease can be accounted for by averaging over single airports with catchment area inside of the affected region. Fig. 5.8 shows real world (**A**) and hypothetical (**B**) outbreak scenarios and risk posed to worldwide, European and US locations. Airports inside affected regions were not included in the ranking as import risk has no meaning when an outbreak is already ongoing in the area. The figure confirms observation from profiles: outbreaks pose high threat to airports in their indirect geographical proximity. For example multiple Indian airpots are at risk in case of MERS outbreak, South-East Asian airports are endangered when outbreak happens in East Asia. Two outbreaks pose disproportionally high threat to USA compared to other regions of the world: ZIKA and a hypothetical outbreak in the Caribbean. In both cases all but one of top 10 worldwide threatened airports are part of the United States. Multiple airports listed in top 20 worldwide ranking from Fig. 5.5 appear in this two scenarios. Compared to scope ranking, the subset of airports is strikingly different for all outbreak scenarios. For all cases top gate airports are distributed over several geographical regions often showing one or few dominant gates for the outbreak. Best example of this is Ebola outbreak. In case of import risk there is no single airport which is at disproportionally high risk. Interestingly in case of Ebola outbreak two biggest gates, London Heathrow and Paris Charles de Gaule, are neither in top 10 worldwide nor in top 7 Europe wide threatened airports. Once again this underlines that gate airports starring high scope are at relatively low risk of initial disease importation. In every case similar set of airports inside Europe is endangered. The picture is slightly more diverse inside the US, but multiple hubs appear in each bar chart. In cases of Zika and a Caribbean outbreak the subset of threatened airports is different from other depicted scenarios.

Figure 5.9 shows a direct comparison between scope and import risk profiles (**A**) and rankings (**B**). As described above, both metrics produce distinct sets of targets. Biggest gate airports are rarely at high import risk. Highest threat of case imports is most often presented to the smaller hubs, which have considerable population size, but do not act as transit nodes. This is an important relationship, which can be easily overlooked, and can lead to underestimation of risk for second order neighbours of the outbreak location.

We must keep in mind that worldwide spreading is dominated by air traffic, but local spreading occurs along different routes. In countries with poorly developed infrastructure local spreading can still resemble wave like pattern as known from historical outbreaks. When the infrastructure is present and commuting is a common practise the spreading can happen faster and wave patterns will be obscured. Locations close to outbreak source will be affected even if import risk calculated based on WAN suggests otherwise. Hence over time affected region will become bigger as the disease spreads locally. Same logic can be applied to

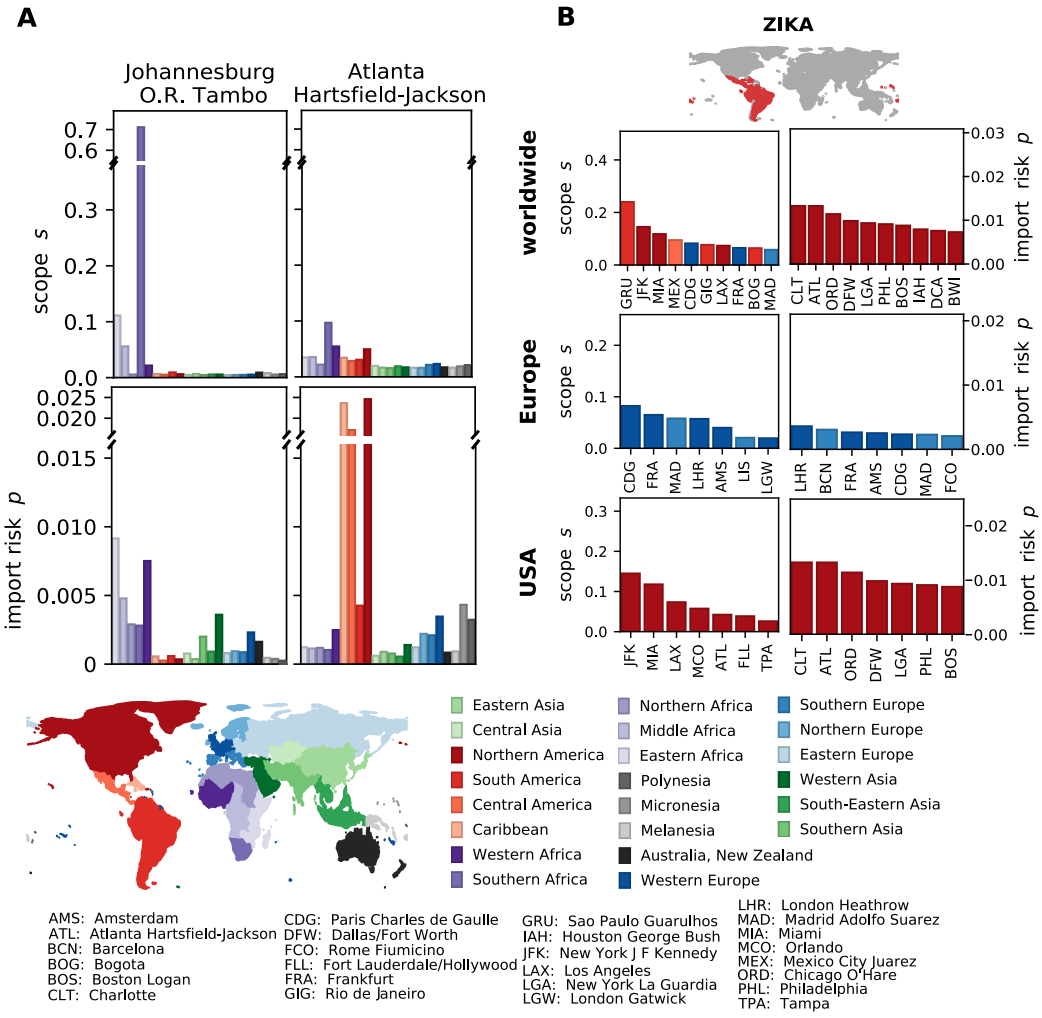

Figure 5.9: Comparison between the scope and import risk profiles (**A**) and rankings (**B**).In both cases the regions and airports indicated by the metrics are different. (**A**) While Johannesburg O.R. Tambo is a major gate in case of an outbreak in Southern Africa, the highest importation risk to it is posed by a different African region. Atlanta Hartsfield-Jackson is not acting as a gate for any of the regions. Nonetheless it is at high risk of case importation from all American regions. (**B**) In the worldwide ranking according to scope multiple European airports appear. While acting as spreaders they experience a low importation risk. In Europe- and US-wide rankings according to either metric deviate considerably, with multiple airports appearing in one ranking exclusively.

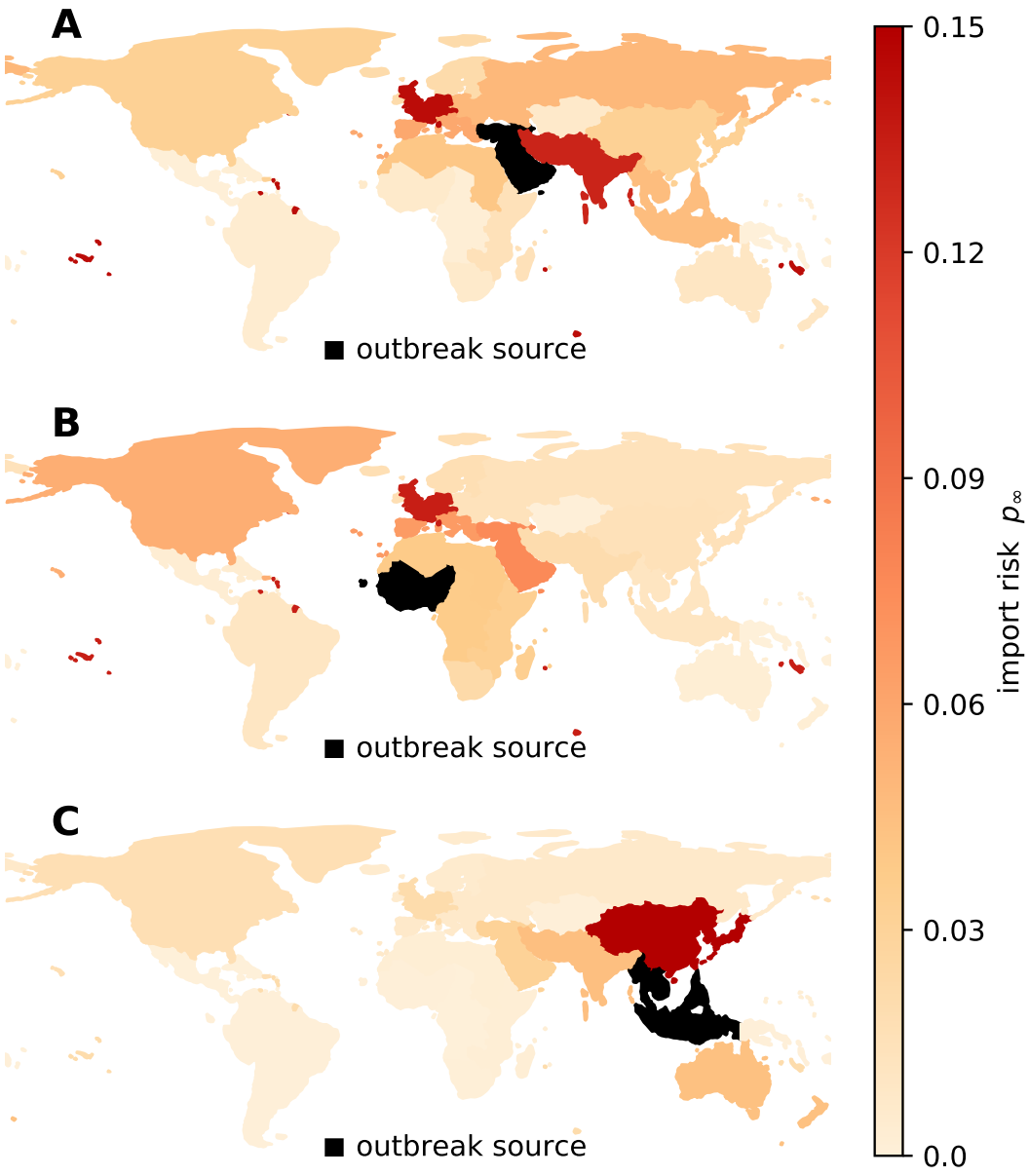

Figure 5.10: Aggregated, averaged import probability in case of a region wide pandemic. Outbreak region is shown in black, no import risk is calculated for it. **A** shows an outbreak in the region of West Asia, **B** - West Africa, **C** - South-Eastern Asia. The colour represents import risk aggregated for an entire region. This view shows that some outbreaks exhibit local patterns, endangering mostly regions in close proximity (**C**). Others pose a risk to remote locations with which they share historical ties (**B**). In both of this cases big parts of the world remain unaffected. Western Asia is region hosting base airpots of multiple big airlines and therefore hubs of considerable size. In case of an outbreak in this region import risk is distributes more evenly across the world.

import location. Even if the destination has a weak coupling to WAN and hence low import risk, disease can be imported into a hub geographically close to the target. From there it will spread through commuter traffic and affect its geographic proximity. To account for this effect it is useful to study relations between broader regions. Fig 5.10 shows how an outbreak in one of the 21 geographical region affects the remaining map. Depicted outbreaks exhibit different types of patterns. An outbreak in South-East Asia (**C**) has a high risk of exporting disease inside own continent, to Asia and Oceania while presenting almost no risk to Europe or Africa. This local spreading pattern is the one captured most by current political thinking when it comes to outbreaks, where treat is evaluated in terms of geographical distance. In an outbreak scenario as depicted in **B** this reasoning can lead to dangerous false estimations of risk when geographical distance is considered. While being geographically closest to its own continent, West Africa poses highest threat to West Europe. Its historical ties with France and Great Britain are reflected in the structure of air traffic network as increased travel fluxes between the countries. This outbreak scenario shows that the influence of history runs much deeper than previously assumed and can shape seemingly unrelated phenomena. In panel **A** a region with high amount of big transit hubs with low population is infected. West Asia harbours multiple base airports of major flight companies connecting Europe and Asia and other destinations around the world. Thus it poses a more uniform threat to all regions including North America, North Africa, Europe and Asia. Once again West Europe is at a highest threat, followed by South Asia. This scenario contradicts to conclusions which would be drawn using geographic distance. As was emphasised in the introduction thinking in terms of geographic distance is not applicable in current age of fast global traffic. Fig. 5.10 shows what differences arise between import risk and geographic distance.

Within the context of local political counselling it is more important to know which locations affects a specific country instead of knowing how risk is distributed worldwide. Fig. 5.11 demonstrates this application for three international airports in different regions of the world. Bar charts show top 10 airports, countries and regions posing a threat to the airport of interest. All airports which appear in the bar charts are medium hubs often sharing the same or neighbouring region with the airport of interest. This is in line with observations made with prior examples. An exception is Puerto Plata which poses a high import risk to London Heathrow (LHR). It can be attributed to colonial and political ties of Martinique, which also appears in top 10 of countries posing highest threat to LHR. Several African countries, which also share historical relations with Great Britain, appear in top 10 for London Heathrow. At the same time no single African airport is represented in the ranking. Airport wide ranking is dominated by european locations, while country ranking is much more diverse in terms of geographical region. In terms of regions London Heathrow shows a strong association with entire Africa, Northern Africa above all others. Again no single airport or country from that region appears in the rankings. This is surprising but possible if medium risk is posed by all locations inside this region, which leads to high import risk over all. Dubai airport shows strong association with African region in terms of airports, countries and overall regional import risk. Only two non african airports appear in top 10: London Heathrow and Sylhet airport. One word of caution is appropriate in case of Dubai Intl.. Assumption that the flux of the airport is representative of its catchment area might be validated and thus the probability of a passenger exiting there is likely to be overestimated. While accumulating high flux Dubai Intl. is mainly a transit airport, but this does not render any of conclusions invalid. When validated this assumption affects the probability from every outbreak location to same extent therefore leaving the sequence intact. For Beijing Intl. airport most airports posing the highest threat are from the same region. The rest of the airports belonging to top 10 are from Oceania as well as multiple countries from the ranking. Oceania is also the region posing highest threat to to Beijing airport. Interestingly for all depicted airports high-

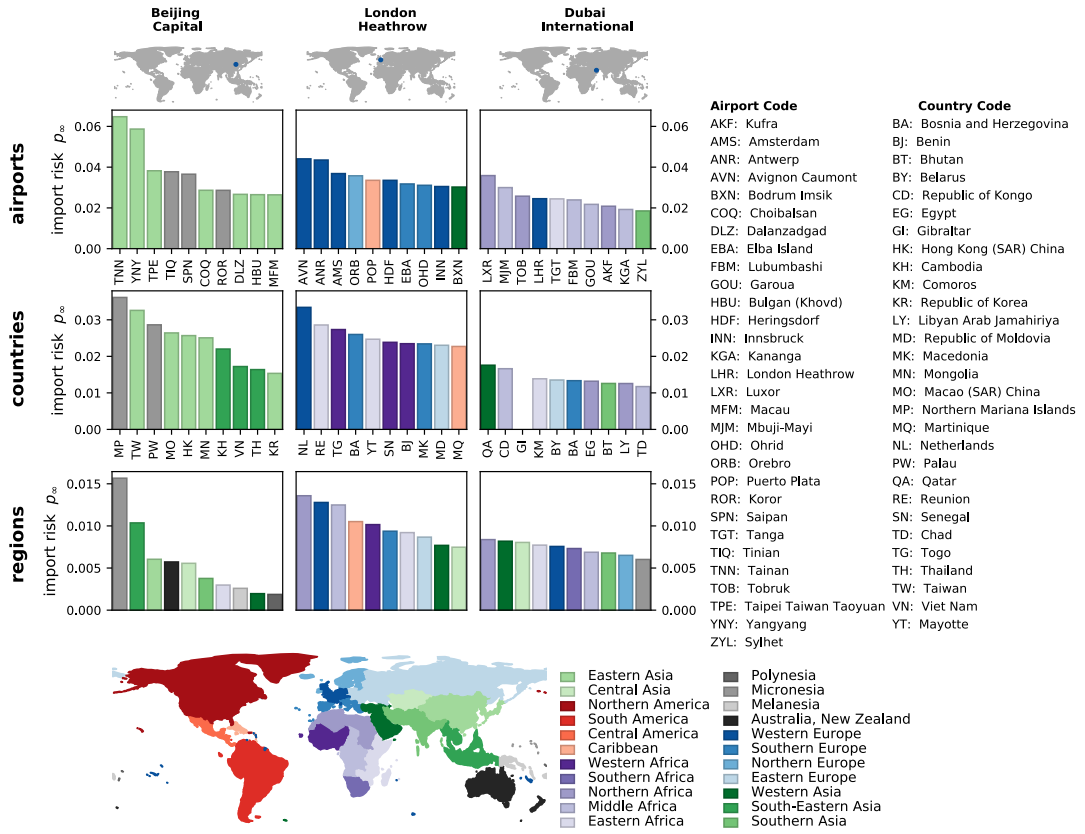

Figure 5.11: Outbreak locations which pose highest threat from the perspective of an airport. Airports lying in the same country as the airport of interest are excluded. Top row shows scenarios in which only the location of the airport is affected, the middle row represents an outbreak in a single country, the bottom row represents an outbreak in a geographic region. Bars are coloured according to the geographic region. The airports are labeled according to the IATA 3 letter code, country labels are ISO two letter codes. This perspective can be used to judge the threat current outbreak is posing to the location of interest. In case of Beijing airport most dangerous locations lie inside the same geographic region. Interestingly the highest threat region wise is posed by Micronesia. From the perspective of London Heathrow airports from different geographical regions pose a threat. The presence of african countries and regions is especially striking. The outbreaks in multiple north african locations pose a high threat for Dubai. In none of the cases is the own region the highest threat. One possible explanation is that the biggest hub of the region acts as a gateway to the rest of the world, therefore taking a role of a transit airport with low likelihood of passengers exiting there. Import risk using air transportation network is not a fitting metric to estimate spreading on the local scale, where air traffic is not the dominating transport.

est threat region-wise is posed by a different region than their own, though we need to be cautious with this conclusion in case of China. In this calculation airports belonging to the same country as the airport of interest were excluded. In case of China this is a substantial fraction of Eastern Asia. An interesting detail is that there is no overlap between locations which grant an airport high scope and pose a threat in means of disease import. This holds when outbreak origins are airports, countries and regions likewise. Once again this observation supports the claim that major gates are at lower importation risk than their children. Other centrality measures did not make this difference between the role a node plays and the threat posed to a it.

5.3 Discussion

The methodology introduced in this chapter provides a scientific way to judge import risk during and prior to pandemic situations. We have demonstrated the application on past, ongoing and hypothetical pandemics. We identified locations exposed to highest threat of case importation, which are not obvious in terms of their geographical location. Though in some pandemic situations geographical reasoning and predictions made by import risk coincide, most of the time very different pictures arise. As we have observed in case of West Africa sometimes history influences seemingly unrelated processes. Hence reliance on geographical information only ignores multiple important factors. Import risk is not considering those directly, but it relies on air transportation network on which the spreading is taking place, thus all influential factors are already contained in the network.

Import risk of a node is highly context sensitive and makes predictions tailored to a specific outbreak location. It is an advantage of this metric over conventional centrality measures which assign a node same value regardless of outbreak origin and spacial spread. While incorporating some specific information about the outbreak import risk does not rely on parameters that are difficult to estimate or prone to variations, e.g. basic reproduction rate R_0. Our approach can be applied on early stages of the epidemic while complicated ODE models, which need proper parametrisation, are available only after several weeks of outbreak. This is especially true for newly emerging or highly adaptable diseases. As described in chapter 3 spreading speed can be estimated using effective trees and can complement import risk.

Presented results support our claim that transit airports even in close proximity of the outbreak are not the most endangered locations. Exit probability of transit hubs is very low, hence reducing the import probability. Nonetheless we have to note that the model is based on an assumption which can be validated by some diseases. In the definition of import risk we require infected agent to terminate the trip and exit at a node to spark an epidemic. During the stay in an airport interactions are limited to a small number of people who might or might not belong to the population of respective node. When released into population inside the catchment area of an airport infected is in contact with native individuals and is more likely to generate secondary cases relevant for the location. For some highly contagious, airborne infections like measles sufficient number of relevant infections can occur inside the airport. Due to high R_0 in a fully susceptible population and airborne transmission route an epidemic can be sparked this way. Nevertheless we consider described scenario to be an exception rater that the rule, either due to low R_0 of most diseases or sufficient vaccination coverage in the population.

Realistic outbreaks are rarely confined to a single location, as they spread over time when containment efforts are delayed or unsuccessful. Hence to account for this scenarios aggregation and averaging procedures are necessary. We have proposed a method of averaging import risk over multiple outbreak locations and account for population sizes at outbreak

sites. Note that with growing size and often geographical spread of the outbreak the threat of it for other locations is increasing. Using our proposed averaging method does not account for this as has been explained in this chapter. Other methods are needed to estimate the propensity or the amount of infected released by a region. In case when origin of the outbreak is uncertain our averaging method is very fit to account for this uncertainty. Apart from this we have demonstrated that aggregation over multiple target locations is also possible. With provided tools it is possible to account for most realistic outbreak scenarios and also make decisions when information is lacking and true outbreak origin is uncertain.

Knowing the import risk still leaves one important question unanswered: what is to be done about it? In the next chapter we explore this question by devising a game theoretical model. At the base of it lies a transportation network analogous to the WAN where nodes are players able to commit resources for epidemic prevention. In contrary to previous chapters we use simplified generic networks for the study to understand basic laws and dependencies which might be heavily influenced by topology. General rules are derived from the model giving general advise and widening understanding of this complex problem.

Chapter 6

The good, the bad and the optimal: resource allocation strategies during emerging pandemics

In this chapter we investigate a model for optimal deployment of mitigation resources in a network of interacting countries. We implement two versions of the model: selfish, with nodes minimising own cost, and pro-social, with nodes minimising global cost. In both cases the node can exercise a limited amount of resources among all nodes of the network to mitigate an outbreak. At each node costs are a combination of invested resources and effective susceptibility to case importations. The mathematical definition of the model presented in this chapter employs game theoretical methods and iterative optimisation. Therefore, after mathematical definition, we explain the specifics of the implementation of the model. We show, analytically and numerically, that purely selfish and cooperative actions do not differ considerably in a single outbreak scenario. In both cases the resources are donated to the outbreak location. However, in a scenario with multiple outbreak locations we find that resource allocation can follow more complex patterns and nodes can fall back on egocentric resource allocations. Furthermore, the amount of resources available to the system changes the final distribution of resources, with wealthy systems allocating more resources to the affected nodes. We show that restricting the nodes from specific strategies gives rise to well known intervention patterns, i.e. ring vaccination, which has proven effective in the field.

Since the onset of the twenty first century multiple pandemics emerged, went global and disappeared after making headlines and claiming lives. SARS was the first epidemic of the century and a wake up call highlighting the speed of spreading and unpredictable jumps a disease can exhibit in a modern world. While it was by far not the deadliest epidemic claiming 774 lives from total 8096 infected, it forced WHO to issue a global alert and contributed to implementation of International Health Regulations (IHR) two years later [109]. The latter agreement, signed by all members of the organisation, was a clear sign that international oversight and coordination was wished for and needed to handle global issues. One of the new tools enabled by the agreement is the declaration of Public Health Emergency of International Concern (PHEIC) followed up by recommendations of actions to contain the threat. Until now PHEIC was issued four times, during the flu pandemic in 2009, polio resurgence in 2014, Ebola outbreak of the same year and most recently in connection with Zika outbreak

and its proven link with multiple neural disorders. Three out of four pandemics occurred in developing countries with low gross domestic product and health infrastructure unprepared to handle outbreaks of this scale. One of the most alarming and distressing pictures arose during the Ebola pandemic which lasted for almost two years with local infrastructure completely overwhelmed at an early stage. In the course of pandemic 28 616 people were infected of whom 11 310 died[114]. This number is believed to be highly underreported for all affected countries as rumours and fear of public stigmatisation prevented many from reporting Ebola cases. Even those who did often could not be accommodated in hospitals and hence were not tested or accounted for in the statistics.

Despite global oversight through WHO international response was heavily criticised by multiple organisations and assessment boards [35, 78, 68, 76, 56]. It was called inappropriate on many levels, underfunded, delayed and badly coordinated. Similar, but less harsh evaluations were made after H1N1 pandemic in 2009 [75]. All of the listed assessments emphasised the role of global engagement and the need of international cooperation in case of large scale pandemics to achieve timely and adequate response. Achieving the latter is a challenge as often interests of countries conflict with global interest. Hence an egoistic country can refuse to comply to WHOs recommendations, free ride by not contributing to the funding or enforce politics which intervene with the course of global action. During the Ebola epidemic this behaviour was displayed by about a third of the IHR countries, among others by Australia and Canada, who were the only high income countries to impose travel restrictions in spite of WHO's recommendation not to do so [82, 90]. IHR lacks any tools of enforcement or punishment of non compliance hence such behaviour is risk free. This phenomenon is a well studied topic in game theory and social sciences. It is investigated in form of simplified scenarios called public good games posing similar dilemma as outlined above. Empirical and modelling studies show that cooperation can only be sustained when means of punishment for non-cooperation are given [30]. Without punishment defectors benefit from their cooperative peers by exploiting them, which leads to decline and disappearance of cooperators in the community. Further it has been shown that if punishment is allowed for all members of the community anti-social punishment can lead to cooperation decline as well [88]. Hence when considered as public good game we must conclude that no cooperation will arise without strong enforcement of the recommended strategy and this frail state can be destroyed by anti-social responses from the participants.

As some countries put emphasis on favouring intrastate matters before interstate relations we expect it to be challenging to implement policies proposed by agreements like IHR. Given this situation it is important to study behaviour and decision making of participating countries when no top-down regulations are imposed or can not be enforced. In context of worldwide pandemics surprisingly little research is done on this. Many publications concentrate on the optimal resource distribution between two or multiple connected meta-populations while assuming that the resources will be distributed by an overarching institution [9, 20, 93, 87, 64]. In reality resources are owned by countries and hence are managed by its government, which is mainly interested in promoting countries interests. While WHO and IHR have resources at their expense they heavily rely on donations, either voluntary or issued by agreements. But again, as no enforcement tools are available contributions can be issued with great delay or refused altogether. Hence we want to study this system as an optimisation problem on a network of selfish agents. As outlined above the resources are owned by the agents and are used to serve self-interest. Nodes are optimising own costs, thus an optimum for one node can be conflicting with best solutions for others. We assume that pro-social behaviour can occur in a pandemic situation thus we incorporate it as a second mode of optimisation. In this case nodes do not optimise costs inflicted to themselves exclusively but minimise network wide costs. We are not interested in emergence of cooperation as our framework is not

designed for it. We want to observe the final distribution of resources across the network given all agents are selfish or pro-social. Of particular interest is the difference in allocated resources between the two optimisation scenarios.

There are multiple challenges connected to this problem which we will not address in this thesis. While there are multiple ways to evaluate a threat of an ongoing outbreak there is no unified framework or gold standard. We proposed another, in our view superior method in previous chapter by defining a probabilistic, network based measure to estimate import risk. Nevertheless we can not expect all countries to rely on the same metric, hence we expect the perception of risk to vary across countries. In the work at hand we will use proposed import risk as a basis for research presented in this chapter and assume that all nodes evaluate the risk the same way. Another issue which will not be addressed is free-riding. It is a widely discussed phenomenon in game theory and in context of vaccination of individuals. While possible in our scenario we will refrain from studying it at this stage. Presented work is aimed at basic understanding of the system which in itself poses multiple challenges. Further we disregard that in real world some countries are tied by historical and political relationships and further interests which alter their cost function and readiness to aid each other. Such values are hard to estimate and are at risk of overcomplicating and obscuring the picture.

6.1 Model

As outlined in the introduction we want to define a system as a network of players, which are optimising cost individually and without coordination. The model is comprised of two main components: backbone based on network theory and the optimisation layer inspired by game theoretical approaches. In our model nodes represent countries and links are connections between those. Note that a connection must enable transmission of disease across this link, e.g infected individuals travelling across a link. A natural example of such are traffic connections, e.g.as described by the world aviation network (WAN). For vector or food borne diseases trade networks can be used. The strength of the link is determined by its flux, in the case of WAN the number of plane seats on scheduled commercial flights. Though we pick the links to be capable of spreading, explicit modelling of the spread is not part of our model. The outbreak is represented by one or many outbreak sources chosen at system initiation. Infected nodes harbour certain amount of infected, who inflict costs. Nodes can assess the probability that an infected will leave the source, travel to and exit at the respective node. Each arriving undetected infected is connected with cost, e.g. for delayed quarantine, extended treatment, contact tracing and and treatment of secondary cases. Thus using the latter and probability of arrival expected cost from an outbreak can be evaluated for each node. These cost can be modulated through donation of resources for disease containment. All nodes possess a limited amount of resources $\rho_n \geq 0$, which can be used for disease prevention. These represent the amount of money, personnel or supplies a country n can mobilise in a case of a pandemic. We do not specify the form in which resources are available to keep the model as simple as possible. In real world amount of resources will vary across countries and is likely to be correlated with flux. We want to study general phenomena so we do not require this proportionality in our model, but we investigate it computationally. A node is allowed to distribute its resources over a network by splitting them between nodes, donate all resources to one node or refrain from spending any. All actions can be combined or performed with arbitrary fractions of maximal available resources. To reflect how likely infected at node n are missed we introduce a property called susceptibility Q_n. Susceptibility Q_n indicates the probability that an infected will be released from or arrive at node n undetected, hence $0 \leq Q_n \leq 1$. Susceptibility of a node can be reduced by donations and possibly be

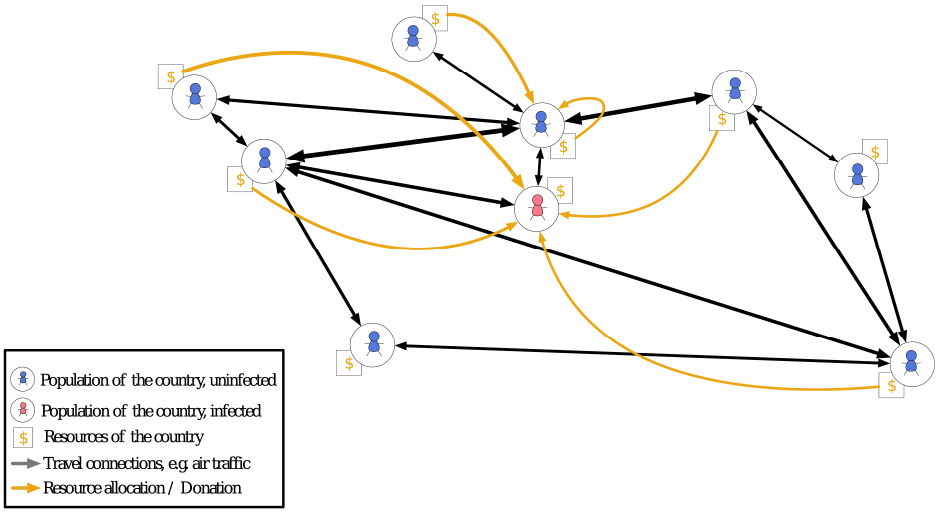

Figure 6.1: Schematic explanation of resource allocation model. The basis of the model is a network of players, i.e. countries, where a link f_{nm}, coloured black, represents a traffic connection from node n to node m (e.g. number of passenger in the air transportation network). Nodes have two model specific properties: their infection state (red pictogram - infected, blue - healthy), susceptibility Q_n and maximal resources ρ_n. The state of a node indicates if the node is a source of an outbreak and is hence releasing infected agents. Susceptibility of a node is a measure of how easily an infected traveller can traverse or enter the node. When an infected enters a node it causes costs for the node, e.g. to quarantine infected, trace contacts etc. A node n can donate all or a fraction of its resources to any node in the network, itself included, to combat disease spreading (investment represented by orange arrows). Susceptibility of the node, which receives resources for disease prevention is decreased. When $Q_i = 0$, i no infected agent can traverse of enter the node. Hence if susceptibility of source is decreased to 0 outbreak is contained. Nodes in the model are players optimising a cost function. The cost function consists of two parts: 1. resources spent for disease prevention and 2. cost of importation weighted by the probability of import cases. Depending on the mode of the model a node is either optimising only own costs (selfish mode) or cost of the entire network (pro-social mode). No explicit disease spreading or cheating behaviour is implemented in this model. The emphasis lies in the resource distribution across the network in equilibrium given a certain state of the network. Depicted investment is not an outcome of a simulation.

pushed down to 0. In real world reduction of susceptibility can be achieved by exit or enter screening in airports, where infected can be quarantined fast and efficient without inflicting high costs. When $Q_n = 0$ is achieved no infected can exit at node n or proceed to travel. Hence this model allows a complete prevention of the epidemic: if the susceptibility of the source is reduced to 0 pandemic is successfully brought under control.

We emphasise that proposed system is not equivalent to a public goods game. It resembles public good games formulated to model vaccination behaviour with an important distinction. While the cost function of each node has a similar form, it is numerically different. Network topology determines a unique import probability for each node, thus making the agents distinct. Accordingly each node can have its own optimal strategy. The latter is in contrast to the solution of classical public good games. Important implication of the multi-agent approach, well known in game theory, is that actions of agents change the state of the system and affect decisions of their peers. Hence action of a player, which was optimal before action of its peers needs to be revised afterwards. This way optimisation has to be performed iteratively until the system reaches its final equilibrium. Note that in this model resources donated to a node can be withdrawn at the next step, hence representing verbal commitment rather than physical allocation. Iterative nature of this process makes both analytical solution and computations more challenging.

We further emphasise that, while having similar motivation, our model is distinct from the model presented in [103]. Wang et. al. restrict the network to two interconnected nodes and consider the final number of cases for cost estimation. While this allows for elegant analytical solution, this system is difficult to tract as soon as the number of nodes increases. As our emphasis lies on exploring the problem in a multi-node system and on varying network topologies we do not tract the final outbreak size. In the publication by Wang et. al. nodes are symmetrical in terms of their import probability. We on the other hand pursue a system which gives rise to different subsets of nodes, close and far from the source of infection. Further we do not require nodes to spend all available resources, which is the case in [103]. Thus we present a more general system, which is solved with different tools.

In following we will introduce exact mathematical notation for the system and explore it. Multiple strategic decisions can be solved analytically on general topology, even though final equilibrium can not be determined regardless of the simplicity of system components. We explore the equilibrium computationally on different network topologies, outbreak scenarios and parameter combinations.

6.1.1 Mathematic definition

The base of the model is a network with N nodes and L links between those. Links are routes with flux F which bear spreading potential . We refrain from a detailed derivation of the network from data as it has been described in Chapter 3. Each node has a maximal amount of available resources ρ_n, which can be distributed across all nodes of the network. We call this distribution a strategy of the node n, r_n, where $(r_n)_k$ is the amount of resources n donated to k and $0 \leq \sum_k (r_n)_k \leq \rho_n$. Note that the latter implies that not all resources have to be spent and that the resources are strictly positive. A matrix $R$with $R_{kn} = (r_n)_k$describes investment state of the entire network. Further we rely on the notion of probability that an agent travels along a link from n to k, P_{kn}, and import risk $p_\infty(n|i)$ as defined in equation 5.1. In this model import risk can be modulated by allocation of resources, which changes susceptibility of a node $0 \leq Q_n \leq 1$. Susceptibility of a node fulfils all requirements to be

treated as probability and thus can be easily incorporated into import risk

$$p_\infty(n|i,Q) = Q_n q_n \sum_{t=1}^{\infty} (\widetilde{\mathbf{S}}^t)_{ni}$$

where

$$\tilde{\mathbf{S}}_{ni} = P_{ni} Q_i (1 - q_i)$$

Note that if $Q_n = 1$ for all n we return to previous definition of import risk. For a single path $\omega = \{1, ..., L\}$ new import risk is $p(\omega) = Q_L q_L \prod_{k=1}^{L} S_{kk-1}$. If $Q_n = 0$ for any $n \in \omega$ no disease can be transmitted along this path. If the susceptibility of the source of the outbreak $Q_i = 0$ disease can not be transmitted to any nodes, thus the outbreak is fully prevented. Note that in the new definition import risk is strictly smaller than defined in Eq. 5.1, hence

$$\sum_m p_\infty(m|n,Q) \leq 1$$

As shown in previous chapters we can apply the same metric when there are more than one outbreak origin. Averaging and aggregating can be implemented the same way as described for initial definition of import risk in equation 5.2.

There are multiple ways how to translate donated resources into susceptibility reduction. Simplest assumption is to imply linear dependency of a form

$$\begin{cases} Q_n(R) = 1 - \sum_m R_{nm} & for \sum_m R_{nm} \leq 1 \\ Q_n(R) = 0 & else \end{cases} \tag{6.1}$$

This definition is in agreement with $0 \leq Q_n \leq 1$. Linear function of Q_n has the benefit of being analytically tractable, further it is a definition with fewest assumptions and is hence favoured by the law of parsimony. In the work at hand this definition of susceptibility is used unless stated otherwise. Two further functions are plausible: power law and sigmoidal. Power law function account for saturation or threshold-like phenomena, depending on the exponent. It is defined in the same intervals as the linear version with exception of $Q_n(R) = 1 - (\sum_n R_{nm})^{\varepsilon}$. Exponents ϵ above 1 reflect that arbitrary small donations are of little use, full potential can be harnessed only above a certain investment threshold. Exponents lower than 1 reflect that some residual risk remains even if high amounts of resources has been donated. Both of these effects are simultaneously accounted for when sigmoidal function is used. We will explore all three possibilities numerically in following sections.

We have defined import risk as a metric to assess the threat posed to a node by a specific outbreak location. Import risk indicates how likely it is that an infected leaving the outbreak location will arrive at the node of interest. While irrelevant in prior chapter we need to define how many infected are exported by the node and how many remain inside. This can be rarely estimated from data and has to be assumed. We thus define the number of infected individuals in the source node i, λ_i. The fraction of infected who leaves node i is Γ_i. For simplicity we assume that $\Gamma_i = \Gamma$ for all i. At the target node each individual inflicts certain cost for its isolation, treatment and contact tracing, C_I^0. Using the above we can define expected cost inflicted by the outbreak at node i to node n

$$C_O(n,R) = \left(\Gamma p_\infty(n|i) + (1-\Gamma) Q_n \delta_{ni}\right) C_I$$

where

$$C_I = C_I^0 \alpha$$

is the global cost of the outbreak in case with no interventions, α reflects severity of the outbreak. If the outbreak is distributed across multiple origins Y import risk is calculated across all of those as described above

$$C_O(n,R) = \left(\Gamma p_\infty(n|Y) + \sum_{i\in Y}(1-\Gamma)Q_n\delta_{ni}\right) C_I$$

Cost inflicted by the outbreak can be reduced by investment as outlined above. Investment is associated with loss of resources, hence it must be incorporated into the full cost function which is as follows

$$\begin{aligned} C_n(R) &= \sum_k R_{kn} + C_O(n) \\ &= \sum_k R_{kn} + \left(\Gamma p_\infty(n|i) + (1-\Gamma)Q_n\delta_{ni}\right) C_I \end{aligned} \tag{6.2}$$

with $\delta_{ni} = 1$ if $n = i$ and 0 otherwise.

In selfish mode function 6.2 is minimised by each node disregarding the global cost. While general form of the function remains the same, import probability varies strongly depending on position of a node. Calculation of import probability is the reason why exact analytical solutions for general topologies are not feasible. Optimisation of cost must be done iteratively, which is the second challenge when it comes to the approach at hand, computationally and analytically. After one node has decided upon its strategy remaining nodes of the network decide upon theirs. When every node in the network has updated its strategy the process is repeated.

In pro-social mode cost inflicted to the entire network is minimised by each node

$$C_n(R) = \sum_k R_{kn} + \sum_k C_O(k,R)$$

Note that investment of other nodes is not entering the equation as it can not be influenced by the acting node and will only introduce a constant factor to cost of every strategy. To decide which strategy to adopt cost for all possible strategies need to be calculated. Strategy with minimal cost is chosen. For this reason we don't need to be concerned with a constant factor added to the cost.

To gain some basic understanding of the model we consider some simplified scenarios analytically. We do not aim to calculate final resource distribution analytically as the problem at hand is too complex to allow it. We rather seek to understand some aspects of the decision making and prioritisation of investment recipients. To make our questions analytically tractable we need to impose some restrictions on the model. Those apply only to the analytical solutions and will not be posed on computational model. Until the end of this subsection we consider strategies in which only one entry of the investment vector r_n is changed at a time. Further the change $\Delta(r_n)_k$ is small and $\sum_m (r_n)_m + \Delta(r)_k \leq \rho_n$, meaning node n owns enough resources to increase investment without affecting the rest of its strategy. Beyond that we require a linear relation between susceptibility Q_n and investment, $Q_n = 1 - \sum_m R_{nm}$. While in most general case loops are allowed when paths in a network are considered it is plausible to assume that agents using air traffic network avoid unnecessary flights. Thus a path traversing same node multiple node is unlikely. This implies that an agent leaving the source of infection will not return to it after a mere pleasure flight using a different node as transit. Hence $p_\infty(i|i) = 0$ and the cost function for infection source simplifies to $C_i = \sum_k R_{ki} + (1-\Gamma)Q_i$. Under this circumstances infected only has two strategies:

invest in self or refrain from using any resources for disease prevention. Thus below we will not discuss the case of acting node $n = i$ as the solution for the source node is given above.

We consider a scenario in which a selfishly acting node n is comparing two strategies r_n and r_n^* associated with cost $C_n(R)$ and $C_n^*(R^*)$. We can rewrite $C_n(R)$ as $C(r_n)$ as investment of nodes $k \neq n$ remains unchanged during this calculation. We impose that both strategies differ in investment in one node $(r_n^*)_k = (r_n)_k + \Delta(r_n)_k$ and remain equal for all others $(r_n)_m = (r_n^*)_m, m \neq n$. Hence $C_n^*(r_n^*) = C_n(r_n + \Delta(r_n)_k)$. We can use Taylor expansion to calculate this:

$$C(r_n^*) = C(r_n) + \Delta(r_n)_k - C_I\Gamma\frac{\sum_{\omega\in\Omega} p(\omega)}{Q_k}\Delta(r_n)_k$$

with $\sum_m R_{nm} \in [0,1]$ and thus $Q_n \in [0,1]$, $\Omega = \Omega_k$ is a subset of paths which lead from infected to acting node and traverse node k. For detailed derivation see Appendix 8.4. Change of strategy imposes two sources of cost change: difference in invested resources and change of import probability. In calculations below we define that $\Delta R_{kn} > 0$.

Using derived formula we can compare discrete strategic decisions. Assume node n is comparing two strategies as outlined above. Additionally $C_n(r_n^*) < C_n(r_n)$, thus r_n^* is a better strategy. Using the formula derived by Taylor expansion we can calculate the necessary conditions. For the sake of clarity only final results will be presented here, for more detailed derivation see section 8.5. Condition for r_n^* to be superior is as follows:

$$\frac{Q_k}{C_I\Gamma} < \sum_{\omega\in\Omega} p(\omega)$$

Multiple insights can be generated from the above formula. Strategy r_n^* is more likely to be beneficial when cost of disease import is high or high fraction of infected is leaving the source. This lies in line with the expectations that no interventions are needed when disease is mild or if infectious population is confined at the source. Further if $Q_k > C_I\Gamma$ investment in k is never a beneficial strategy, which means that if Q_k is pushed to a low, but non-zero value further investment is disadvantageous. Hence $C_I\Gamma < 1$ provides a general threshold below which no investment will be observed among non-source nodes. Further higher probability of case importation also makes strategy r_n^* more beneficial. Note that susceptibility of recipient of investment Q_k and import probability $\sum_{\omega\in\Omega} p(\omega)$ are not independent and the latter can be rewritten as $Q_kQ_nq_n\sum_{\omega\in\Omega}\prod_{m=1}^{L} S_{mm-1}\prod_{m\in\omega;m\neq k} Q_m$ and hence Q_k can be eliminated from the equation. Thus exact value of Q_k at the moment of decision making does not influence decision making. Note that above is valid inside the boundaries $Q_k \in [0,1]$ and under the assumption of linear function of Q_k.

Another important facet to understand is the prioritisation of nodes when it comes to investment. Using same tools as described above we can derive condition for node b to be a better fit for investment than node a. Assume node n has a strategy r_n, from which it considers to deviate. There are two alternative strategies r_n' and r_n^* which differ from the default by $\Delta(r_n)_a$. Strategy r_n' directs more investment to node a, thus $(r_n')_a = (r_n)_a + \Delta(r_n)_a$ and $(r_n')_m = (r_n)_m$ for $m \neq a$. In strategy r_n^* additional investment is donated to node b, thus $(r_n^*)_b = (r_n)_b + \Delta(r_n)_b$ and $(r_n^*)_m = (r_n)_m$ for $m \neq b$. Further $C_n(r_n') > C_n(r_n^*)$, meaning that strategy r_n^* is the preferable strategy. For simplicity we assume that $\Delta(r_n)_a = \Delta(r_n)_b$. Using Taylor expansion we can derive conditions for above to be true.

$$\frac{\sum_{\omega\in\Omega(a)} p(\omega)}{Q_a} < \frac{\sum_{\omega\in\Omega(b)} p(\omega)}{Q_b} \tag{6.3}$$

where $\Omega(a)$ and $\Omega(b)$ are sets containing all transmission paths from source to acting node which traverse nodes a and node b respectively. Thus investment decision is based solely on import probabilities and susceptibilities of the nodes. If the susceptibility of nodes is equal, best strategy is to invest in the node which set of paths accumulates higher import probability. There are two nodes which lie in every transmission path, source node i and acting node n. Hence $\Omega(i) = \Omega(n) = \Omega$. As $\sum_{\omega\in\Omega} p(\omega) \geq \sum_{\omega\in\Omega(k)} p(\omega)$ for all k, i and n are always among the best ones to invest in. If import probabilities via the nodes are equal node with the lower susceptibility is preferred target. Consequently once a node received investment it becomes a self-reinforcing process. Taken together these observations suggest that infected node is the best candidate for investment. Source node is present in every transmission path of every node, hence it is viable candidate for every player in the network. Acting node itself is likely to be among investment candidates for himself only. Thus source of infection has the capability to accumulate more donations than any other node and hence cause its peers to follow the same strategy. While we can not prove this case on a general topology we can show how probability for this to happen unfolds in the simple cases.

It is important to point out that it is possible for another node to be as good as source or acting node, but assuming that all susceptibilities are equal no node can be a better candidate. If the source is connected to the network via one link, its sole neighbour k is participating in every path, thus $\Omega(k) = \Omega$. In this scenario this node is as likely to receive investment as the source of investment. This outcome was observed in our simulations.

To study how probability of receiving investment changes with the position of the node we consider first two steps of optimisation. At each time step a node on the network is chosen with uniform probability. This node evaluates the situation and chooses the optimal strategy to adopt. Constrains imposed above remain, meaning that only one entry of the strategy vector can be changed at a time, susceptibility function is linear according to definition 6.1. We further assume that at initiation of the model all nodes have the same susceptibility. For further calculation we combine this assumption with the insights from above paragraph: only nodes which are present in all transmission relevant paths are considered. For exact derivation of formulas used in figure 6.2 see Appendix 8.6

First we look at investment on a chain of length L where source is placed at index 0. Chain topology means that there is only one path, $|\Omega| = 1$ and a node n_k is present in transmission path for all nodes with index larger than k. We can calculate the probability of each node on the chain to receive investment in step $t = 0$ and $t = 1$. As can be seen in figure 6.2 **A** probability declines with the index of the node. Hence first node on the chain has the highest probability to receive investment. First node is a viable option for all nodes on the chain, while a node which lies further down the chain will not receive any investment from nodes with smaller index. Once a node with low index $k \ll L$ has received investment at $t = 0$, all nodes with index $m > k$ will invest in this node at $t = 1$. If $k = 0$ any next chosen node will dedicate its investment to source node. In this case equilibrium state is determined at the first step of the simulation. Nonetheless there is non negligible probability that nodes with higher index will receive investment and the network will settle in a different state.

A more general example is presented in fig. 6.2 **B**. We consider a network which consists of two components $\mathcal{N}_1$ and $\mathcal{N}_2$. Node i is the source of disease and thus present in each transmission relevant path. Node j connects components $\mathcal{N}_1$ and $\mathcal{N}_2$ and is present in each transmission relevant path for nodes in component $\mathcal{N}_2$. Thus i has a higher probability of receiving investment as there are more nodes for which it is a viable choice. We can compare probabilities of i and j receiving investment at $t = 0$ and $t = 1$ depending on the size of components. When $\mathcal{N}_1$ is big compared to $\mathcal{N}_2$, i has considerable advantage over j. As $\mathcal{N}_2$ increases this advantage is shrinking, but i retains considerably higher probability of receiving investment even when $\mathcal{N}_2$ is ten times bigger than the first component. If at $t = 0$ i received investment

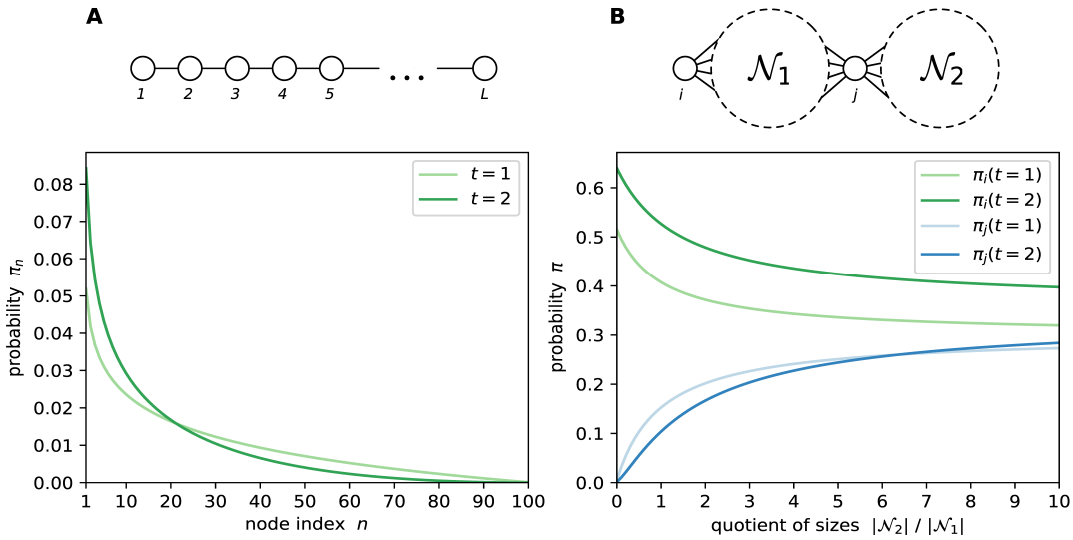

Figure 6.2: Probability to receive investment with regard to position of receiving node on a chain (**A**) and a more general topology (**B**). Lighter shade represents probability at $t = 0$, darker shade at $t = 1$. In all cases we assume that susceptibilities of nodes are equal at $t = 0$. Plot **A** shows probability distribution on chain with length $L = 100$. Probability steeply declines with index of the node, being close to 0 for terminal nodes. Probability to receive investment is highest for node with index 0. Decline is steeper for $t = 1$, which is a carry over from $t = 0$: in case if node at index 0 received investment at $t = 0$ it is guaranteed to receive investment in all subsequent steps. **B** shows the probability distribution for two exemplary nodes on a general topology. Depicted network consists of two components $\mathcal{N}_1$ and $\mathcal{N}_2$, x-axis shows the quotient between sizes of those. Node i is the source of infection and is present in every transmission relevant path to nodes in both network components. Node j lies in all transmission paths from source to nodes in $\mathcal{N}_2$, but is not a valuable target for investment of nodes in $\mathcal{N}_1$. If the second compartment of the network is negligibly small compared to the first probability of i receiving investment heavily outweighs probability of investment in j. Even if $\mathcal{N}_2$ is ten times of $\mathcal{N}_1$there is still a considerable gap between nodes j and i. This becomes even more profound for $t = 1$.

each subsequently chosen node will invest in i. Hence, similar to the chain, final equilibrium is decided at first step. This is reflected by higher probability of i to receive investment at $t = 1$.

Both cases presented in figure 6.2 show that source node is expected to receive considerable amount of donations. This fortifies previous analytical results that source of infection is an attractive target for resource allocation. It also shows that early investment in the source can determine the final state of the model. While presented analytical results covered special cases we were able to gain important insights in the behaviour of the model. We have clear indications that source of infection is a good candidate for investment and is likely to be the node to receive all investment. Note that we have only considered outbreaks with one source of infection. In the implementation of the model we can restrict this assumptions and study more complicated scenarios.

6.1.2 Implementation

To observe final strategies of nodes a numerical solution of the model is necessary. Thus a model as described in the previous section was implemented. We run simulations on different network topologies: square lattice, Erdős–Rényi and networks based on Delaunay triangulation. In the latter case nodes were positioned at random in 2d-space and links were introduced according to Delaunay triangulation rules. In all cases fluxes over each link were set to be equal, i.e. $f_{ij} = 1$ if there is a link connecting i and j and 0 otherwise. Thus $P_{mn} = 1/k_n$ with k_n being degree of node n. Further all nodes were initiated with $Q_n = 1$ and available resources according to ρ_n.

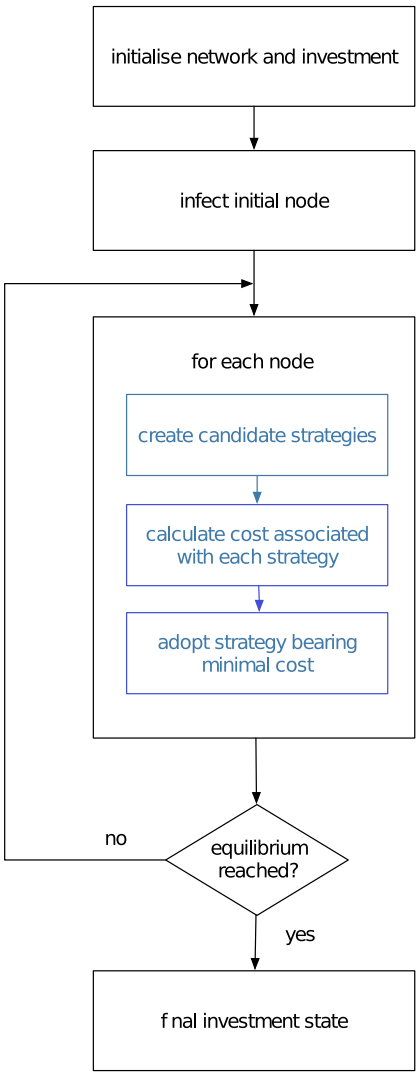

Figure 6.3: Schematic representation of model's implementation. After initial network set up strategies of nodes are optimised iteratively until system wide equilibrium is reached. Latter is the case when no node has changed its strategy over certain number of iterations. We ensure that in course of one iteration strategies of all nodes have been updated.

After initial network setup outbreak node was chosen and its state was set to 'infected'. For each node a set of paths for possible infection import was calculated. To make it computationally feasible some restrictions had to be imposed on import risk calculation. In the most general case amount of paths connecting two nodes in a network is infinite, new path can be created by introducing loops. It is plausible to assume that agents

travelling via air transport will avoid unnecessary flights and never traverse a node more than once. Thus we consider only paths without loops. Still remaining number of paths is very high causing long computations times. Therefore all paths with probability $p(\omega) < 4 * 10^{-4}$ were not considered during the calculation. For details on the implementation of the filtering algorithm and choice of the threshold see 8.8.

After network setup is finished, strategies of nodes are updated in multiple iterative rounds. In each round strategies of all nodes on the network are updated. Possible update orders are random, where sequence of nodes is generated randomly, and hazard-based, where sequence is determined by import probability allowing highly threatened nodes to update the strategy first. To update own strategy acting node evaluates the cost of 100 strategies generated according to a modified evolutionary strategy algorithm. Strategy bearing minimal cost is adopted. For more details on the procedure see Appendix 8.8.1. Next node according to the order rule is chosen to update its strategy. Note that after one round every node has updated its strategy once, ensuring that all strategies are updated frequently. Rounds are repeated until system reaches its final equilibrium. System was considered as equilibrated when no change happened for sufficient amount of rounds.

It must be pointed out that in this implementation strategy of infected is trivial. By disallowing paths with loops no importation back into the source was possible, hence for outbreak origin only valuable strategy is to invest in self. We consider this behaviour natural as an affected country will likely dedicate available resources to combat emergency which is already happening inside of it.

In contrary to analytical solution multiple functions for susceptibility were tested computationally. Implemented functions are shown in figure 6.4. Different functions represent different challenges that can be faced when deploying countermeasures. Linear function represents the simplest case where donation in directly translated into susceptibility decline. While it is the simplest case it assumes no limitations on the receiving end. Power law function of a form $Q_n = 1 - x^{\varepsilon}$with $x = \sum_k R_{nk}$ accounts for threshold and saturation phenomena. If exponent ε is greater than one initial investment leads to slow susceptibility decline. Only after certain amount of resources is accumulated susceptibility drops substantially. This scenario reflects that arbitrary small amount of resources can not be used effectively. If the exponent of the function is below one, susceptibility decreases very fast initially. At low susceptibility disproportional amount of resources must be invested to push it further below. This reflects that residual risk is difficult to eliminate, e.g. preventing asymptomatic cases from travelling can be connected with unreasonable amount of resources. Sigmoidal function reflects both limitations and can be thus considered as most realistic case. All functions can be tuned by choice of parameters. Parameter estimation in this context is close to impossible, hence we use power law and sigmoidal function to investigate general change introduced by these functions. In all other cases we use the simplest function. Consequently linear function is used if not stated otherwise.

Our model was implemented in Python and Julia. It is based on multiple third party packages available through public repositories. Julia code used for the work at hand will be made available on git repository. We established default values for all parameters and used those if not stated otherwise. For full list see Appendix 8.9.3. All simulations were performed on networks with 100 nodes and average degree of 4 when possible. The model was tested using different network topologies, with very similar results. Networks were initiated wth $Q_n = 1$ and $\rho_n = \rho$ for all n. If not stated otherwise full system has sufficient resources to fully prevent the spreading by reducing susceptibility of the source to $Q_i = 0$. The latter is not guaranteed as resources suffice only if a significant fraction of nodes decides to participate in investment. Further we allow some excess of resources to verify that donations were stopped when $Q_i = 0$.

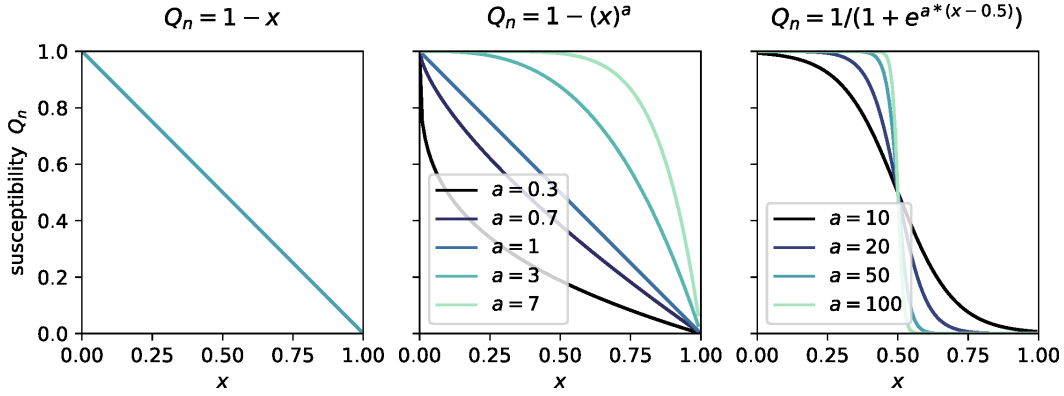

Figure 6.4: Implemented susceptibility functions: linear, power law and sigmoidal. In the graph $x = \sum_n R_{mn}$. Different functions account for different phenomena. Linear function represents the simplest case where any amount of resources is directly translated into benefit. Power law function accounts for two effects depending on the exponent a. If $a > 1$ arbitrary small amount of investment will have no benefit, meaning that an initial threshold has to be overcome before decline in susceptibility becomes steep. If $a < 1$ susceptibility drops very fast but large amount of resources has to be used to conquer residual risk. Sigmoidal function accounts for both, initial lag and slow decline at low susceptibility values.

6.2 Simulated results

In sections above we defined the model in mathematical terms and gained insights about several special cases. Due to the complexity of the formulated model numerical simulations are needed to investigate the final equilibrium. Using implementation described above we conduct extensive simulations to scan parameter space and study how imposed restrictions and different network topologies affect the final state. Both modes of the model were evaluated: pro-social and selfish. While the outcome of the model differs across the range of some parameters it is not sensitive to the choice of network topology. Main determinant of strategy of a node is the import probability from each source present in the network. Import probability distribution depends on network under investigation, but network topology in itself did not add any additional value to the analysis. Networks with profound small world properties more often result in similar distances to multiple infected. This bears specific implications which will be discussed below.

As has been suggested by analytical solution, when outbreak is seeded by one source it is likely that source of infection will receive all donated resources. This was confirmed by numerous simulations across a wide range of parameters in both modes, selfish and pro-social. All nodes, which contributed resources for disease prevention, donated their resources to the source node regardless of the optimisation mode as can be seen in figure 6.5 **A**. Two main differences arose depending on whether agents were selfish or not. First, in pro-social case infected i was fully immunised, meaning $Q_i \approxeq 0$. In a system of selfish agents susceptibility of the infected node was pushed to a low value, but it never reached zero. This is consistent with analytic findings, suggesting that there is a susceptibility threshold below which investment is not beneficial for a selfish agent. Furthermore it indicates similar behaviour as [103], where optimal allocation in selfish case is conflicting with globally optimal solution. Is in the cited paper, agent interested in global good will invest more resources in the affected region compared to a selfish player. Note that this finding is consistent across very distinct

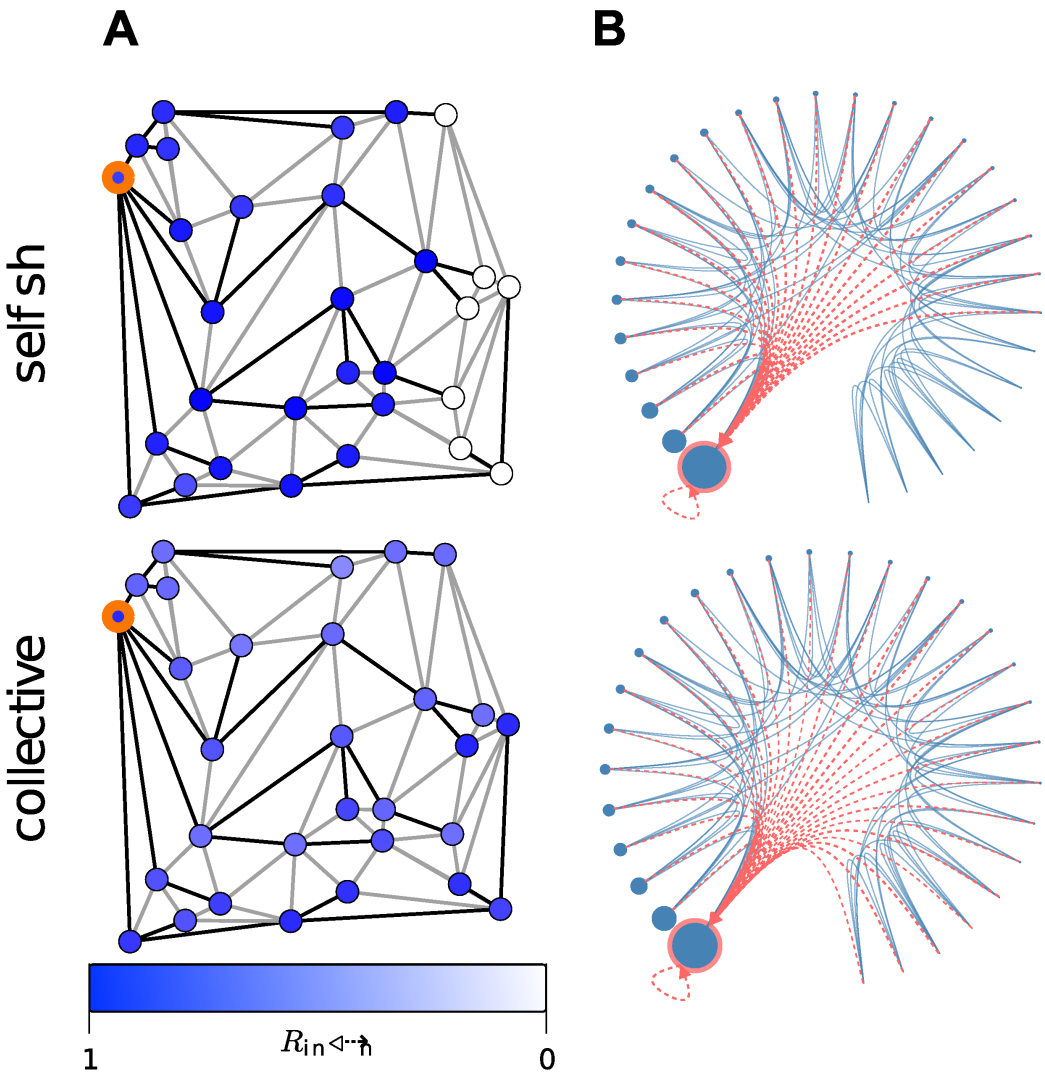

Figure 6.5: Example of resource allocation in a network with one infected. (**A**) Highlighted node is the infected node i, colour intensity indicates the fraction of available resources ρ_n donated by the node to the infected i. Black coloured links are part of the shortest path tree, which is the collection of fastest routes from the infected to all other nodes. Upper row shows optimisation results in selfish , lower row in pro-social mode. Same Delaunay-based graph was used in both scenarios. There is enough resources in the system to fully suppress disease spreading by reducing susceptibility of the infected $Q_i = 0$. In selfish optimisation case nodes invest all available resources if they invest any. In selfish equilibrium $Q_i > 0$ in all conducted simulations. In pro-social optimisation scenario more nodes participate in the investment, but only view invest all available resources. In pro-social equilibrium $Q_i = 0$. (**B**) Same network as shown in **A**, nodes are arranged clockwise with decreasing import risk, red highlighted node is the infected. Node size scales with individual cost at initiation. Blue lines indicate links, red lines indicate investment with the arrow pointing towards the recipient. In selfish scenario there is an import risk based cut off below which investment of resources is not a valuable strategy. In pro-social scenario all nodes participate in investment regardless of import risk.

methodologies used to study the problem.

Second, in selfish case import probability to a node determines the strategy of the respective node. Hence nodes in close proximity of the infected invest close to all available resources. As can be seen in figure 6.5 **B** in pro-social case distance does not play a role. As the cost of the entire network is optimised, it is not dependent on node's position on the network. Also on average each node donates less resources which is compensated by higher participation of nodes in the investment. Note that this is a consequence of optimisation algorithm, where nodes are likely to find suboptimal strategies with lower overall donation amount in the first round, hence donating less than all available resources. Thus other nodes fill the investment gap until full immunisation of the source. Note that this finding is in stark contrast to perceived optimal and implemented reactions to pandemics by single countries. While countries often operate and think in egocentric means our model suggests that this solution is far from optimal even in selfish scenario. Our results go in line with arguments about how the interconnected nature of modern world has changed the patterns and mechanisms of disease spreading and hence requires a different response.

Different susceptibility functions didn't change this general behaviour. When using Power law function with exponent $\varepsilon > 1$ system can settle in no-investment state if initial plateau of the function is too broad. When resources of a single node are less than necessary to breach the plateau no node will commit any resources. Implementing communication or adoption of disadvantageous strategies with low probability can solve this issue. When $\varepsilon < 1$ pro-social systems behaviour remains unchanged, in selfish system final state settles at higher values of Q_i as the payoff of investment decreases with decreasing susceptibility. Sigmoidal susceptibility function combines both of the effects described above.

A more interesting picture arises when number of sources in the network increases. In pro-social case general behaviour is not changed: resources are donated to infected until those are fully immunised. Position of the donor does not influence the choice of the recipient. On contrary strategy of selfish nodes is determined by the import probability. Hence introduction of new competing sources has a big impact on the final state of the system. Figure 6.6 **A** shows how increased number of outbreak sources changes the strategy distribution in final state on a small square lattice. A lattice serves as a good visualisation since importation probability corresponds strongly to the distance from infected in the visualisation. Number of outbreak sources is increased from left to right, infected is marked by a star. Single source realisation underlines what was stated in the previous paragraph. All allocated resources are directed to the infected node. Remote nodes refrain from investment. When more than one infected is present in the network self-investment arises as a dominant strategy for nodes, which are equidistant to the sources and hence have similar import probability from both. We can explain this strategic choice following the argument posed by equation 6.3. In case of two remote outbreak sources node n itself is the only node which is present in all transmission relevant paths. Hence it is likely that $\sum_{\omega\in\Omega(n)} p(\omega) > \sum_{\omega\in\Omega(i)} p(\omega); i \in \mathcal{I}$ where $\mathcal{I}$ is the set containing all infected. Hence even if $Q_n > Q_i; i \in \mathcal{I}$ this does not outweigh the difference in probability. With increasing number of infected number of nodes equidistant to at least two is increasing and so does the fraction of self-investers in the network. This visualisation enables us to observe regions of dominant strategies arising in the network. In close proximity to the infected dominant strategy is to invest in it, in the middle of the network self-investment is the strategy of choice. This demonstrates that egocentric strategy can be the optimal choice, but certain requirements need to be fulfilled. Hence adopting self-investment strategy without proper consideration is likely to be disadvantageous. When observed on networks with less regular structure regions are less obvious in network representation. Main finding remains the same on all types of networks: nodes whose threat is dominated by one source are advised to invest in the respective source, nodes with similar

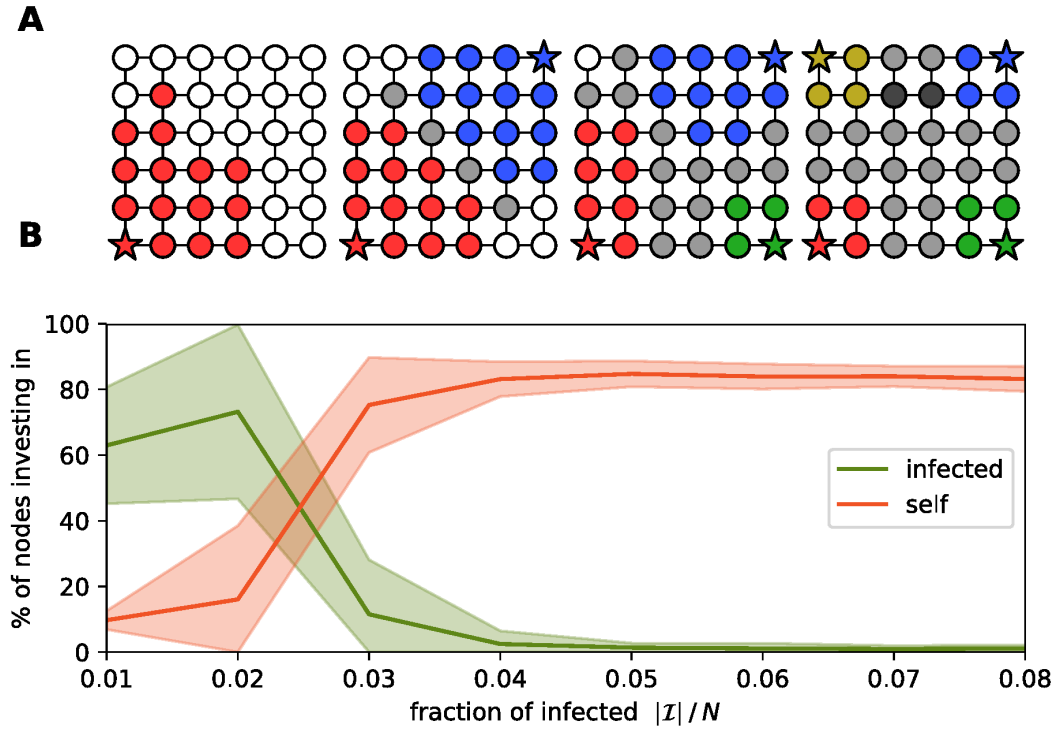

Figure 6.6: Change in strategies with increasing number of infected in selfish mode. (**A**) Change in resource distribution when additional infected are introduced into the network. Infected are marked by stars, colour of the nodes shows investment in infected of the same colour. Grey nodes are self-investers, white nodes do not contribute any resources. Darker grey nodes on the right are following a strategy different from all previously mentioned. Presented strategy distribution presented here is a result of simulation. Nodes in proximity of an infected donate their resources to closest infected while nodes further away either don't invest or invest in self in case of more than one infected. The amount of self-investers is increasing with increasing number of infected. Self-investment is the dominant strategy for nodes which are at same risk from multiple infected. (**B**) Fraction of infected in the network versus the investment strategies. $\mathcal{I}$ is the set containing all infected in the network, N is the number of nodes in the network. Red line shows the fraction of nodes investing in self, red shaded area is one standard deviation from the mean. Green line and shading correspond to nodes investing in infected. Only a small fraction of infected is needed to eliminate investment in the infected completely. Note that front region is subject to high variability which will be discussed on page 90.

import probabilities from multiple infected should self-investment.

Figure 6.6 **B** shows the change in strategy distribution of a selfish system when fraction of infected is increased. Simulations were run on Erdős–Rényi networks with 100 nodes, hence $|\mathcal{I}|/N = 1$ means one infected is in the network. Three percent are already sufficient to dramatically reduce amount of donations to the sources and five percent eliminate investment in the infected fully. This observation has implications for pandemic disease prevention. It highlights that once disease has spread or sufficient crises have accumulated across the network self-investment might remain as the only viable strategy. Hence early alert and timely negotiations about containment efforts can boost the amount of resources infected nodes will receive.

Some parameters which are predetermined in a real world scenario can be manipulated in our model. By varying maximal available resources we can change the wealth of the system. The latter has big influence on the final state of the system in both, pro-social and selfish modes. Figure 6.7 shows aggregated results from simulations on Erdős–Rényi networks in case of two outbreak sources. As outlined above import probability is the determining factor when it comes to strategy of a node, hence in the plot nodes are positioned according to initial import probabilities from infected i and k (x- and y-axis respectively). In each network labels i and k were assigned so that in final state $Q_i < Q_k$. Colour reflects the strategy of a node in final state: green and blue indicate investment in infected i and k respectively and red shows self-investment. Mixed strategies are represented by RGB mixed colours, i.e. strategy of investment in self and infected i corresponds to colour yellow and investment in self and infected k corresponds to violet, with exact hue determined by fractions donated to either node. Nonetheless these strategies are rarely observed, mostly in highly regular networks like lattices. Rows show scenarios with different amount of system wide resources, columns represent different optimisation modes. This plot reiterates the observation that distance to infected does not determine the strategy in pro-social optimisation case. When plenty resources are present in the network, meaning $\sum_n \rho_n \gg 2$ and hence are more than sufficient to immunise both infected, susceptibility of both sources is pushed to zero in pro-social optimisation. No segregation in zones of dominant strategies can be observed. When amount of resources is reduced entire investment is directed to a single infected. It is consistent with equation 6.3. Note that equation 6.3 was subject to multiple assumptions which does not apply to simulations, but findings of the equation remain valid. In case of two competing candidates for investment, the one with lower susceptibility is preferred. It is likely that infected i, which receives the first donation, is the only node with $Q_i < 1$ at $t = 1$. As $\sum_n p_\infty(n|i) = \sum_n p_\infty(n|k)$ node i will be preferred for further investment until its susceptibility is pushed to zero. This finding is particularly interesting in the pro-social context: one infected is being abandoned and all resources are concentrated on one source. From ethical point of view this is a questionable decision, but our model suggest that it is optimal from pure cost perspective. This is consistent with finding reported in [93] and [64], where the question is addressed with a different methodology using two node system and evaluating the cost as final number of infected. Nonetheless the conclusion drawn is that equilibrating the infection load, analogous to immunising both nodes in parallel in our model, is the worst strategy. Rowthorn et.al. find that the best strategy is to invest in the node with lesser infection load and concentrate the resources on this node only. This is well in line with our finding.

In selfish mode division into regions with different dominant strategies is very distinct. Diagonal line shows region of equal import probability from both infected. As stated previously node i receives more investment than node k. It is evident that there is a high fraction of nodes, which are at high threat from i and simultaneously low risk from k. Hence when randomly picking first node to update the strategy it is more likely to choose a node with

strong preference to invest in i. As explained above this reduction in susceptibility increases the likelihood that i will receive further investment. Even when i will not receive investment in the first step it is likely that subsequently chosen nodes still have stronger preference to invest in i as k has only small basin of investment. When resources are scarce nodes in close proximity of k possess too few resources to reduce susceptibility of this source enough to facilitate investment from more remote nodes. Thus for nodes in the middle ground best strategy lies in self-investment. Note that the region of self-investment is shifted towards source k. This implies that certain sources will be at disadvantage due to their position on the network. This can happen when one node is a hub and the second is peripheral. In this case nodes close to peripheral infected are likely to be close to the hub, while there is a substantial fraction of hub's neighbours which is remote from the peripheral source. In real world networks peripheral nodes often indicate developing countries, which suffer from conflicts and have poor economies. Thus they are more reliant on external aid in case of extraordinary events like big pandemics. Our results emphasise that in a selfish environment vulnerable countries are likely to be left behind. Hence an overarching organisation with own resources like WHO, which is acting for a global good, is necessary to balance the difference in donations.

When amount of available resources is increased self-investment is eliminated and investment in the infected becomes only dominant donation strategy. Note that the strategy of nodes in proximity to i remains unchanged as they were already investing in the source. Nodes which were self-investing in the low resources scenario, on the other hand, allocate their resources to k when resource availability is increased. This effect can be attributed to a stronger susceptibility reduction of k through wealthy neighbouring nodes. Its susceptibility is decreased sufficiently to outweigh the difference in import probabilities of path sets and make investment in k dominant strategy. Importantly this behavioural change does not depend on the exact distribution of resources. We tested uniform and random distribution of maximal available resources as well as assignment of maximal resources proportional to node's degree. The outcome is not sensitive to this change and depends solely on the amount of resources present in the entire system as shown in figure 6.8 **B**. Considering this behaviour global non-profit organisations like WHO can facilitate the same effect. Donations into outbreak sources which are not well connected to the rest of the network can prompt self-investing nodes to reconsider their strategy, even when they act selfish. For this strategy to be exploited good knowledge of network structure and import probabilities is needed.

In both selfish scenarios we also observe a distinct region of nodes refraining from any kind of investment. For nodes in this region import probability from either source lies below a certain threshold. Note that there is a linear threshold with respect to each infected, applied independently and a less linear threshold applied to nodes which self-invest. We see this as an indicator that investment in one source does not influence the import probability from the other substantially, hence both processes can be viewed as two separate decisions. Using first derivative of the cost function we can calculate exact threshold for investment in any node m

$$\frac{Q_m}{C_I \Gamma} = \sum_{\omega \in \Omega} p(\omega)$$

where Ω is the set of transmission paths in which m is participating. Assuming that source k participates in negligible fraction of relevant transmission paths from source i and vice versa we can analytically recover the threshold value seen in the plot. Further we can conclude that the threshold for investment in self will be a diagonal in the visualisation we have chosen as target node is present in transmission paths from both infected. For full derivation see Appendix Section 8.7.

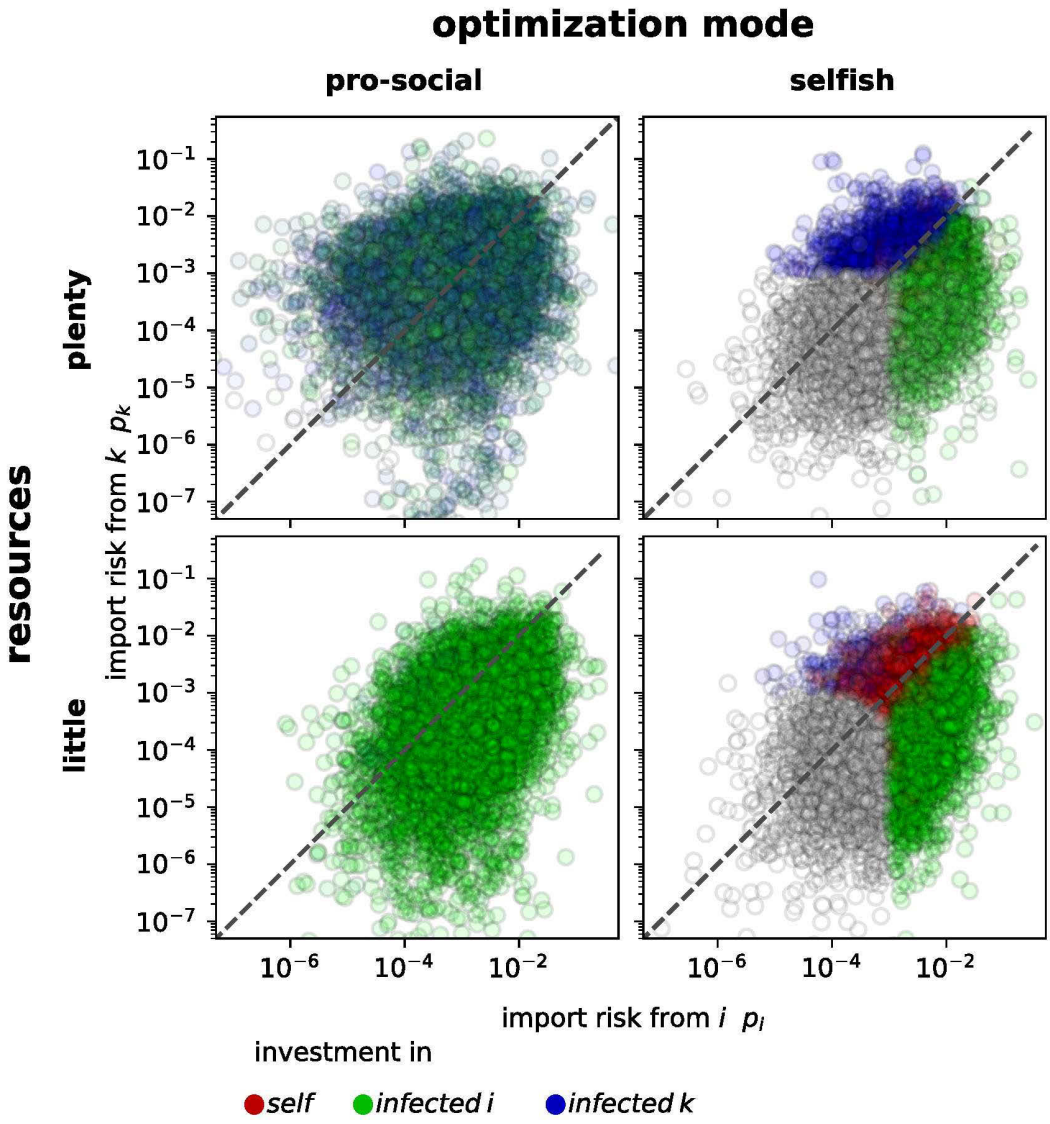

Figure 6.7: Pro-social and selfish optimisation with varying amount of resources. Each dot in the plot is a node from an Erdős–Rényi network (100 networks, 2 infected per network). Infected nodes are not included in the plot. The colour of dots shows the strategy: self-investment (red), investment in infected i (green) or investment in infected k (blue). Mixed strategies are represented by RGB mixed colours. Dashed line shows region of equal import probability from both infected, $p_i = p_k$. The axis show import probability from i an k. Infected are labeled so that $Q_i < Q_k$ in final state. Plenty resources setup:$\sum_n \rho_n \gg 2$, disease can be fully prevented. In this regime all invested resources are distributed between infected in both optimisation modes. In pro-social scenario import risk posed to a node does not play any role for its strategy decision. In selfish optimisation it is the major determining factor. Scarce resources setup:$\sum_n \rho_n < 2$, disease can not be fully prevented. In case of pro-social optimisation all resources are donated to one infected until he is fully immunised ($Q_i = 0$). The residue, if any, is donated to the second infected. When the agents are selfish self-investment arises. In selfish scenario i receives considerably more investment than k. This behaviour is discussed in Fig. 6.9. The linear threshold separating non-investors from nodes investing in the infected is discussed on page 88.

Cost of infection is another parameter, which can influence the amount of donations directed to source of infection. Costly disease leads to increase in donations directed to the source as can be seen in 6.8 **A**. Subplot shows the shift in node fraction investing either in self (red) or in one of the sources (green). Results are drawn from simulations on Erdős–Rényi networks in selfish optimisation mode and two outbreak sources. Other network topologies lead to similar results. Not surprisingly at very low cost of importation any kind of donation is disadvantageous. As import cost increase self-investment is the first strategy to arise. At higher importation cost this strategy is displaced by investment in the source of infection. Costly disease can reduce self-investment in the network, but it is not capable of eliminating it.

Similar picture arises when system-wide resources are varied. While we do not define explicitly what causes cost of importation to increase there are multiple possibilities for this. A disease with high R_0 can be more costly to bring under control after initial importation. Severe progression of disease which requires intense treatment and special quarantine can increase expected cost. Public panic connected with news about import cases can require additional efforts to manage as has been observed in case of Ebola. These findings indicate that increase in perceived severity of the disease should prompt countries to donate more resources to source nodes. We would argue that it is not the observed behaviour and bigger threat leads to higher isolation and self-centred countermeasure deployment. This is an important misconception which has to be cleared to enable efficient pandemic response. Further our model shows that in a poor system self-investment is more prevalent, thus poor world-wide economy will have negative impact on pandemic response regardless of the wealth distribution. It is an important message which our model strongly supports.

Figure 6.8 **C** shows scatter plots explained above for three different system states. In this plots labels k and i were assigned according to node index, hence without regard of investment received. **C1** depicts self-investment dominated regime, which can be observed in system with low amount of resources. Note that this state can not be induced by low importation cost alone. We can further observe that the linear threshold is softened in the area of self investment as is predicted by the analytical result. Self-investment region reaches outwards from the region of equal import probabilities spanning a broad region. When system-wide resources are increased high fraction of self-investment is replaced by investment in one of the infected. In the range of equal import probability from either sources self-investment remains dominant (**C2**). Nonetheless it can be eliminated almost completely by high system-wide availability of resources or by a very costly disease as can be observed in **C3**.

While averaged pictures are conclusive and bear important information we observe high variability among different realisations of networks and source placement while keeping the rest of the parameters constant. In few realisations infected nodes receive close to no investment. In several simulations only one of the two infected receives investment. In majority of cases both infected receive high amount of donations. Figure 6.9 shows final amount of donations granted to both sources across realisations with varying system-wide resources. Each dot represents a single network realisation. Amount of investment needed to decrease the susceptibility of a node to 0is 1. Therefore if $\sum_{n;i\in\mathcal{I}} R_{in} = 2$ both infected are unsusceptible and disease can not spread. As expected from figure 6.8 at low resources infected nodes receive no or very little resources. With increasing resource availability donations into infected increase, with final states stretched out over a wide range. At superfluous resources three distinct final states arise: $\sum_{n;i\in\mathcal{I}} R_{in} = 0$, $\sum_{n;i\in\mathcal{I}} R_{in} = 1$ and $\sum_{n;i\in\mathcal{I}} R_{in} = 2$. When systems are filtered according to their final state as indicated by black boxes in figure 6.9 and plotted we can see a difference in import probability distribution. Scatter plot in figure 6.9 **3** shows probability distribution in networks where both source nodes receive considerable

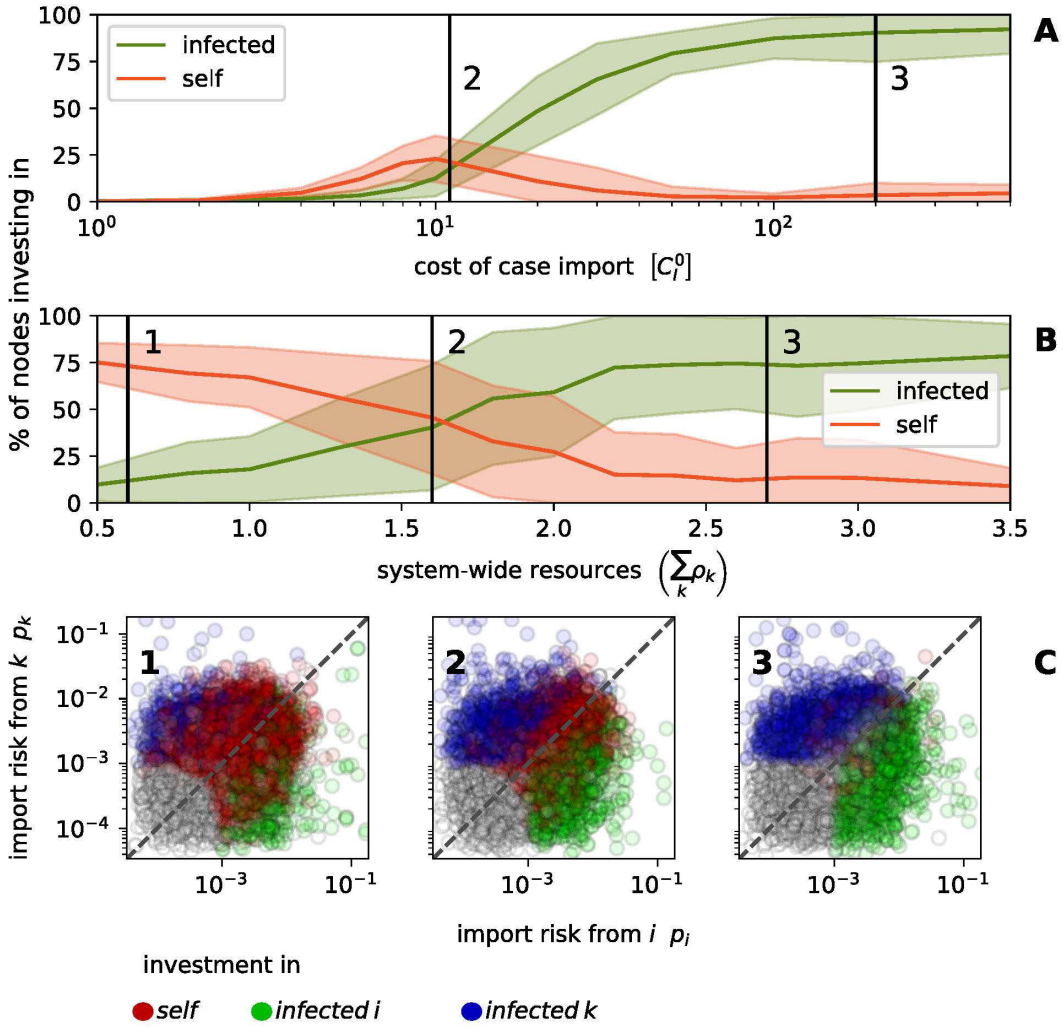

Figure 6.8: Influence of cost of infection (**A**) and system-wide resources (**B**) on strategy distribution within the network. Simulations were conducted on 100 Erdős–Rényi networks with $\langle k \rangle = 4$, $N = 100$ and two infected (i, k) per parameter combination. Red line shows the fraction of nodes investing in self, red shaded area is one standard deviation from the mean. Green line and shading correspond to investment in infected. The reason for high standard deviation is discussed on page 90. **C** shows a scatter plot of 100 realisations of networks using same parameters as indicated by the number in **B** and **A**. Infected nodes are not included in the scatter plot. The colour of \ dot shows the strategy: self-investment (red), investment in infected i (green) or investment in infected k (blue). On the x- and y-axis is the import risk from i and k respectively. Optimisation is selfish in all cases. Increasing cost of disease or amount of resources available to the network promotes investment in infected. The distribution of resources across nodes does not play a role. When the parameters are set as in **B1** or **B2** they are not sufficient to completely eradicate the disease. In scenario **B3** the eradication is possible. Self-investment in all cases arises in the region where import risk from both infected is approximately equal. With growing amount of available resources self-invester region shrinks, leaving nodes exactly at the diagonal last to change their strategy to investment into infected.

amount of investment. It can be seen that there are plenty nodes at high threat from source k but not i and vice versa. We will call this set of nodes investment basin of i and k respectively. As has been discussed previously presence of such nodes with strong incentive to invest into one infected facilitates global investment into source by reducing its susceptibility. Similar effect can be observed in figure 6.9 **2** for source i but not for k. As indicated by blue triangle, nodes in close proximity of k exclusively are missing.

Hence, there is a region of self-investors in broader region around source k. Source i in contrary receives full investment from its neighbours and nodes at higher threat from i than from k. In networks where infected receive no investment as shown in 6.9 **1** all nodes are concentrated around the middle ground with only few nodes in proximity of exclusively one source. Hence we can conclude that source needs an investment basin to initiate investment. If such is lacking region in proximity of such source is likely to self-invest. This implies that two very central nodes as sources can promote self-investment in the entire network. Note that proximity to infected alone is not sufficient to promote investment into a source, it has to be coupled with low import probability from the second infected. There is an exception to this rule which is discussed on page 94. Results discussed until now indicate that there are three viable strategies: invest in self if node is at approximately same risk from multiple infected, invest into infected if node is located in its investment basin or do not invest if import risk is too low. For several reasons there can occur a situation, in which one or multiple of those strategies are unavailable. For example if an outbreak is happening in a country inaccessible for aid for political reasons or deployment of countermeasures is impossible due to a conflict.

To study this we have simulated our model with several strategic restrictions. Bar charts in figure 6.10 **A** represent the fraction of total system-wide resources donated to nodes in different groups: green represents investment in infection sources, blue - direct neighbours of sources (first shell), purple - second order neighbours (second shell), red - self-investers and light grey shows investment different from listed possibilities. Dark grey bar indicates the total amount of donated resources. Note that the height of the total investment bar can be lower than the sum of all remaining bars from the chart. Nodes can belong to multiple categories at the same time hence contributing to multiple bars simultaneously. These simulations were performed on Erdős–Rényi networks. Simulations on a square lattice with otherwise equal parameters are shown in 6.10 **B**. Colour intensity indicates the amount of resources on a logarithmic scale received by the node. Figure 6.10 **1** shows default model regime with no restrictions. In both selfish and pro-social optimisation modes main fraction of investment is directed to infection sources. It is considerably higher in pro-social optimisation mode. In selfish mode minor fraction of self-investment is present with total investment amount lower than in pro-social optimisation. This behaviour can be observed on lattice as well, where both optimisation methods result in same distribution.

In figure 6.10 **2** investment in infected is unavailable as a strategy. Selfish system immediately evades to self-investment, with minor fraction of donations to direct neighbours of the source. When optimised pro-socially investment is shifted to the first shell neighbours and total amount of investment is increased to 100% of available resources. Higher amount of investment is necessary since first shell of the source is comprised of multiple nodes, thus many nodes have to be immunised to achieve full disease containment. When simulated on lattice selfish and pro social outcomes are similar: in both cases high amount of resources is directed to first shell neighbours. This can be caused by low number of direct neighbours in this particular setup. In selfish case some nodes in middle ground of the network receive resources, presumably through self-investment. Note that during selfish optimisation both

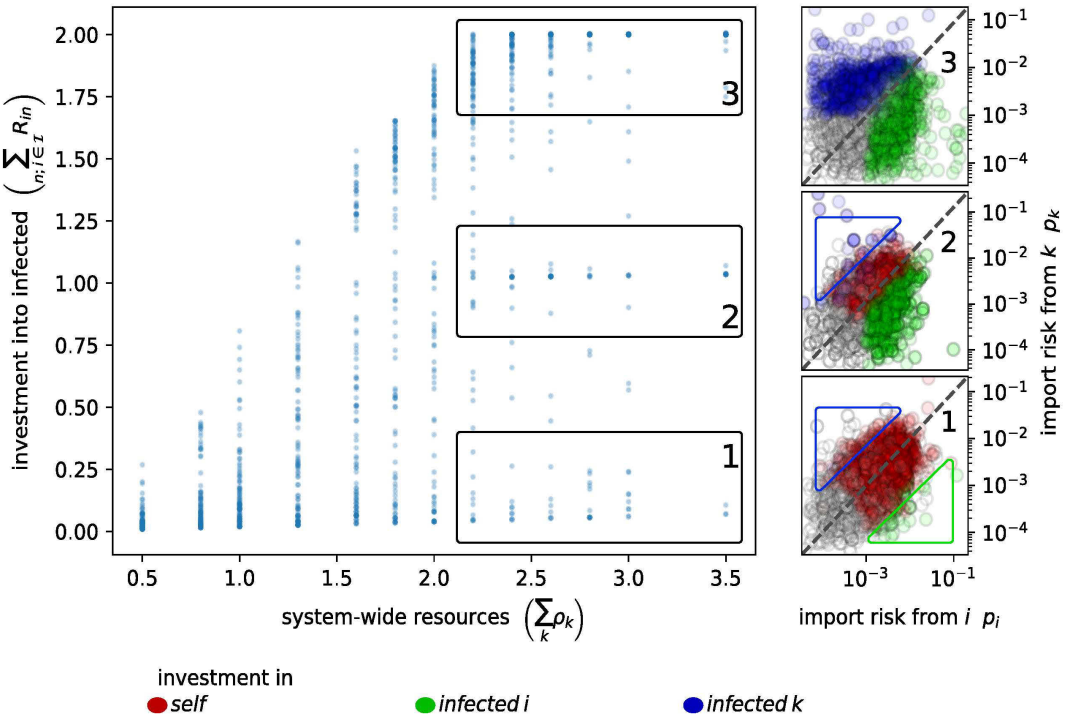

Figure 6.9: Amount of investment into infected in final state of different realisation of Erdős–Rényi networks (left) and strategy distribution with respect to node position (right). **Left:** Each dot represents one network realisation, ρ_k are resources available to a single node, R_{in} is the investment of node n in node $i \in \mathcal{I}$, $\mathcal{I}$ is a set containing both infected of the network. Amount of investment needed to decrease the susceptibility of a single node to 0is 1. Therefore if $\sum_{n;i\in\mathcal{I}} R_{in} = 2$ both infected are fully immunised. At higher values of system-wide resources there are three most abundant equilibria: $\sum_{n;i\in\mathcal{I}} R_{in} = 0$, $\sum_{n;i\in\mathcal{I}} R_{in} = 1$ and $\sum_{n;i\in\mathcal{I}} R_{in} = 2$. Scatterplots on the right are derived from networks in a respective equilibrium (marked by black numbered boxes). **1, 2, 3:** Each dot in the scatter plot is a node from a network. Infected nodes are not included in the plot. Colour of the dot shows the strategy: self-investment (red), investment in infected i (green) or investment in infected k (blue). Dashed line shows $p_i = p_k$. On x- and y-axis are import risk from i and k respectively. Networks in **1** lack nodes that are threatened by infected i, but not k (marked by green triangle) and vice versa (blue triangle), all nodes refrain to self-investment. Networks in **2** lack nodes threatened by k and not i(marked by blue triangle). Nodes for which reverse is true are present, they invest in i. In equilibrium **3** nodes from both sides of the spectrum are in the network, self-investment is adopted by minority. This leads to conclusion that an initial reduction of susceptibility of the infected is needed to make investment in it a beneficial strategy for nodes closer to the diagonal. If this initial seed of a group with strong incentive to invest in infected is lacking nodes prefer to invest in self.

neighbours of source are immunised equally. In pro-social mode one neighbour's susceptibility is pushed to zero while the second receives considerably less resources. This observation is well in line with results shown in figure 6.7, where higher benefit is gained by investing in node with low susceptibility. Note that the concept analogous to ring vaccination arises naturally in our model when investment into infected is not possible, whether due to political, geographical reasons. Our model confirms that ring vaccination is a valid option with low costs. Further our simulation show that this behaviour is beneficial not only in pro-social case but also in selfish scenario. If number of first shell neighbours is low selfish agents utilise immunisation of first shell neighbours, but change the strategy to self-investment if first shell consists of high number of nodes.

Restricting self-investment has no impact on simulations with pro-social optimisation. In selfish mode amount of donations to sources and total amount of investment are increased when self-investment is unavailable. In this case selfish and pro-social investment equilibria show little difference on all topologies. Thus self-investment and investment in the infected are the two best strategies in selfish optimisation case which can substitute each other if one of respective strategies is unavailable. Note that results from pro-social optimisation underlie overall less fluctuation than results in selfish mode. It is not surprising as position of infected nodes play no role for pro-social optimisation and in general cost functions of nodes underlie less variability as they account for entire network. When both, self-investment and investment in infected is restricted in pro-social system nodes of first shell receive all available investment. In Erdős–Rényi networks the outcome is equivalent to scenario with restricted investment in infected only. On lattice a slightly different picture arises. Instead of first shell neighbours a single second shell neighbour is immunised. One possible reason for this is comparatively small size of the lattice used for visualisation purpose, $N = 36$. Both second shell neighbours can not invest their resources in themselves, hence two nodes would have to direct their resources to a different node, which is the second shell neighbour. This node participates in a high fraction of paths and is hence a good target, especially because its susceptibility can be further reduced by first shell neighbours of source.

A different distribution of resources is observed during selfish optimisation: entire investment is directed to first shell neighbours. We attribute this to high incentive of second shell neighbours and nodes on the edge to invest in the first shell. This is also represented in the bar plot derived from Erdős–Rényi networks. Here investment in the second shell is more prevalent, but there is a possibility that nodes which belong to first shell of one source comprise second shell of the other. Overall investment is lower compared to case **2**, when only investment into source is restricted. It is probable, that for some nodes taking the risk of importation of infected is less costly that a suboptimal investment into shells.

In previous simulations we have chosen positions for infection sources randomly. As has been outlined above positioning of nodes with regard to each other can be relevant for the outcome. Further we have made multiple analytical calculations assuming that source k does not contribute considerably to transmission paths from source i and vice versa. In figure 6.11 we purposefully position nodes to violate this assumption. Figure 6.11 **A** shows scenario in which infection sources are not connected by a link. In **B** we ensured that k and i are neighbours, this is also reflected by the tighter cloud in the plot. Two interesting observations arise from these simulations. First, linear border between non-investors and nodes contributing resources to the infected remains unaffected in both scenarios. Calculation of the threshold for i was based on assumption that reduction in susceptibility of k, Q_k, does not change $p_\infty(n|i)$ for acting node n. This assumption is clearly violated when k and i are neighbours, nonetheless the type and value of the threshold remain unchanged. Second, self investment is not observed in simulations where nodes are direct neighbours, even though there is no distinct basin of investment for either infected. It is likely that considerable fraction of paths

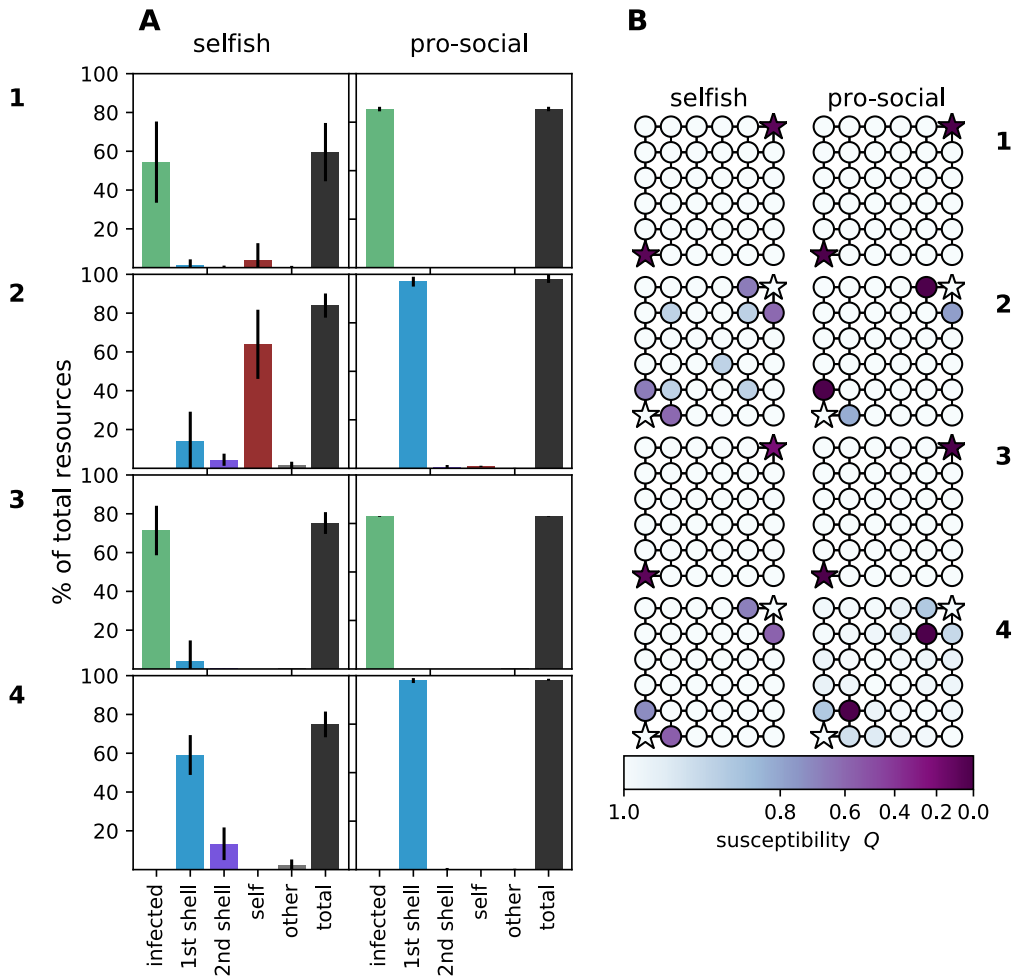

Figure 6.10: Strategy distribution when strategic restrictions are imposed. **1** shows a regular scenario. In **2** investment in the infected is prohibited, in **3** - self-investment, in **4** - investment in self and infected. **A** shows strategy distribution from 100 realisations of Erdős–Rényi networks with size $N = 100$ and two sources. Green bars show investment in source, blue bars - investment in direct neighbours of the infected ("1st shell"), purple bars - second order neighbours ("2nd shell"), red bar - self-investers, light grey - other strategies, dark grey - total investment. **B** shows a simulation on a 6 x 6 lattice using same parameters as respective scenario on the left. Infected nodes are marked by stars. When restricted from investment in source (**2**) selfish system react by increasing fraction of self-investers, investment in 1st shell is minor. In pro-social scenario 1st shell is immunised (analogous to ring vaccination). Due to low number of neighbours of sources lattice results deviate. On lattice 1st shell is immunised in both optimisation modes. In selfish mode some self-investment can be observed, investment granted to 1st shell neighbours is lower than in pro-social mode. Restraining self-investment (**3**) has only minor impact on the final state in both optimisation modes using given parameters. When both strategies are restricted (**4**) selfish system donates considerable amount of resources to 1st shell. Investment in 2nd shell can also be observed, presumably when receiving node is present in a high fraction of paths from both infected. On lattice result differs from outcome in Erdős–Rényi networks. When optimised pro-socially second order neighbour of each infected lying on a diagonal between infected is immunised. This node lies in highest fraction of paths from infected to remaining nodes. Decreasing its susceptibility results in biggest decrease in global risk. As global risk is not important in selfish optimisation this node receives no investment in selfish case.

form i traverses k and vice versa, hence there is additional benefit of investment into each source through import probability reduction from the other. Further overall investment in the network where sources are neighbours is smaller due to interdependency of paths from both sources. Thus to reduce overall import probability less investment into individual nodes is necessary. The results were observerd with two or more neighbouring nodes as sources. These observations have multiple implications with respect to real world scenario. On the one hand when infection spreads from source to its neighbours on the network no sudden jump in self investment will arise. Hence both sources can continue to rely on external aid. On the other hand less resources overall will be contributed to combat the disease. Note that many external factors are not accounted for by the model, e.g. increase in perceived risk through spreading of the disease, which can in turn compensate for observed effect.

Simulation results enable a variety of observations and conclusions about our model. They confirm analytical results suggesting that donation to infection sources and self investment in multi source case are most beneficial strategies if any investment is done. Further they emphasis how model parameters and positioning of infection sources can change the equilibrium state of the model. Simulations also enabled us to show that network topology does not play a dominate role in defining the final distribution of strategies. The latter is mainly determined by values and ratios of import probabilities from all infection sources in the network.

6.3 Discussion

In this chapter we presented a model to study resource allocation problem in context of a pandemic scenario. While remaining very general and simple this model includes a minimal set of features needed to describe the problem and the challenges of decision making. It is grounded in network theory, which proved to be a valuable tool to explore spreading phenomena in general and disease spreading in particular [79, 80, 63, 69]. Second component, based on game theory and optimisation, is ideal to describe decision making processes. Using the model we described a system of interconnected selfish and pro-social agents optimising their strategy in context of a pandemic outbreak.

While analytical findings were only possible by imposing constraints on the model they provided important insight. Complemented by numerical simulations they provided a clear picture about model's behaviour and general governing principles. According to our model main determinant of strategy of a node is the import probability and in case of multiple sources the ratio of import probabilities. There are three dominant strategies: not to invest, invest in one of the sources or in oneself. Latter is only present in multi-source outbreaks and selfish optimisation mode. Which strategy is followed by a node is determined by its position and position of the sources. Considerable differences between selfish and pro-social optimisation are only observed in case of multiple infected. In outbreak scenario with one source equilibrium states of optimisation modes only differ in the amount of donated resources, but not in its recipient. In both cases all donations are directed to the source of infection. This is in sharp contrast to the perceived optimal solution as it was observed during recent outbreaks, where countries chose isolation and ramping up preparedness in their own locale. We argue that our model provides evidence that this reaction is ineffective, even from a selfish perspective. On contrary our model indicates that combined global aid is the most effective strategy to minimise cost connected to imported cases. When multiple infection sources are introduced self-investment emerges as viable strategy for nodes in certain regions of the network. Nevertheless, donation of resources to aid infected remains dominant among nodes in its direct vicinity. During pro-social optimisation source of infection remains

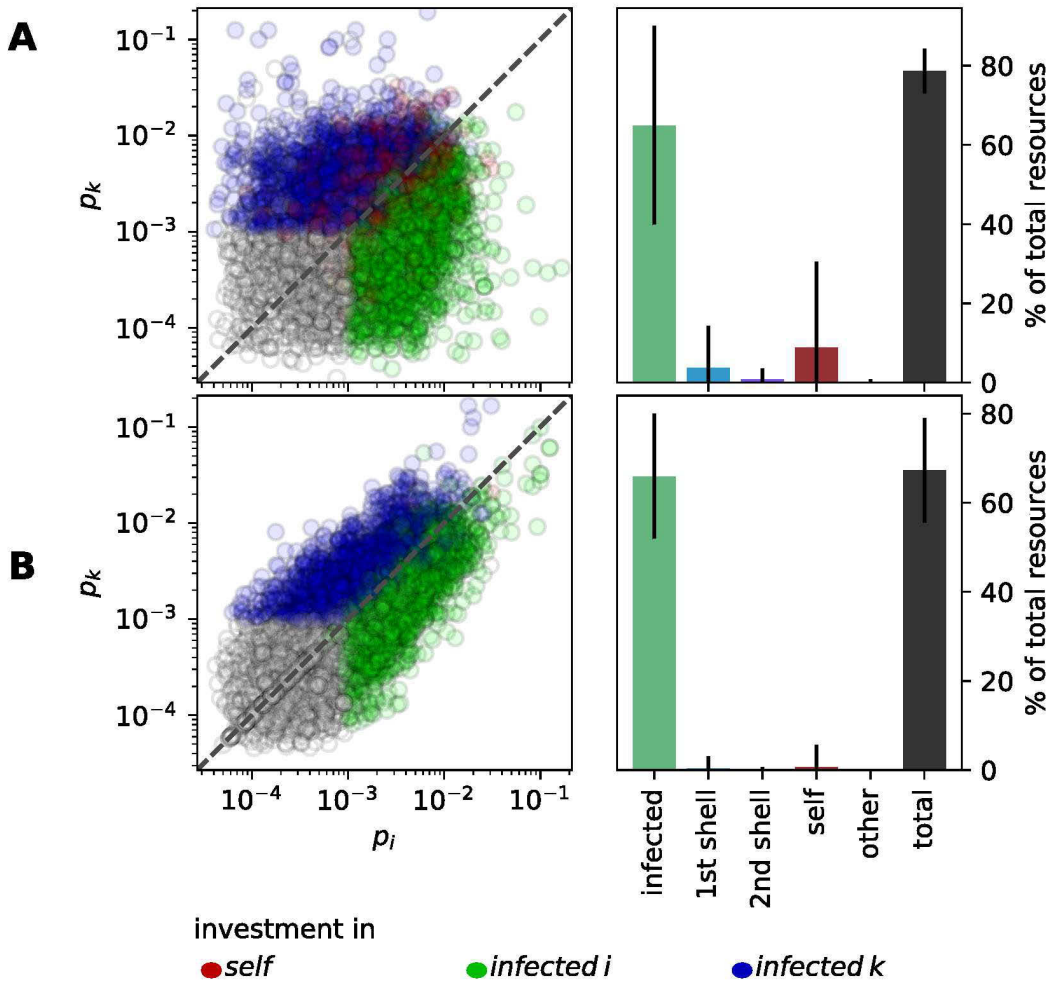

Figure 6.11: Influence of sources' position on equilibrium state in selfish optimisation case. Nodes i and k are infected. In **A** sources are not neighbours, in **B** sources are connected by a link. All parameters are equal. Each dot in the scatter plot is a node from a network, each plot contains results from simulations on 100 realisations of Erdős–Rényi networks ($N = 100$). Infected nodes are not included in the plot. Colour of dots shows strategy: self-investment (red), investment in source i (green) or investment in source k (blue). Dashed line shows $p_i = p_k$. On x- and y-axis are import risk from i and k respectively. When infected are direct neighbours all nodes in the scatterplot are located close to the diagonal $p_i = p_j$, while the nodes in **A** have a higher spread. In contrary to results showed in Fig. 6.9 no self-investment can be observed in scenario **B** even if the nodes are distributed as close to diagonal. In case of neighbouring infected investment into i reduces probability of disease importation from k as it is present in a high fraction of transmission paths from k. Same is true for investment in k. This increases the payoff of investment into either infected. By contrast to **A** self-investment is not beneficial in scenario **B**.

the only viable target for investment.

Further susceptibility of a node is playing an important role in determining how likely a node is to receive investment. Nodes with low susceptibility are favoured for investment, which means that receiving investment further increases the probability to receive additional aid. This process hence exhibits self-reinforcing behaviour. In real world this can be exploited by overarching organisations like WHO to promote desirable outcomes and increase donations in a specific country. By early commitment of resources an organisation can reduce susceptibility of a node prior to decision making of its peers, hence making investment in source an overall attractive strategy.

Apart from these node specific properties some global parameters can also promote certain strategies. Overall resources of the system change the final state of the model, where a wealthy system shows a higher tendency to invest in the infected. This implies that raising global wealth will result in more commitment to global disease control. At the same time this indicates that financial crises and failing economies can start a spiral, where lack of aid in case of severe outbreaks will leads to more failing economies. Note that throughout our simulations distribution of resources played no role. This means that any failing economy will result in decrease of overall commitment and increase the likelihood of self-investment dominated final state. A counteracting effect is that diseases bearing high cost of import cases promote investment in infection sources. Hence our model suggests that severe outbreaks should generate more donations than mild diseases. Regrettably this behaviour is rarely observed in case of real outbreaks. Outbreaks, that are perceived more dangerous lead to isolation politics and egocentric thinking. We hope that our model can be used to challenge this reactions. Intuitively, reduction in export rate Γ leads to lower risk and decline in investment participation. This change can be driven externally by air traffic disruption and reduction of flights or by disease itself. A fast onset and severe symptoms can prevent infected individuals from travelling, thus changing the export rate.

We also demonstrated that position of nodes relative to each other and on the network plays an important role. Network topology influences the distribution of import risk across nodes, e.g. networks with pronounced small world properties contribute to shorter distances, hubs can route infected to high number of nodes but they also introduce considerable drop in import probability for each neighbour individually. Chain-like structures and bottlenecks can divert investment from infected and lead to alternative final resource distributions. When disease sources are introduced independently certain combinations can promote self-investment among subgroup of nodes. Sources lacking neighbours with high import probability from source and low threat from other infected, lack persistent donors. Hence they rarely get high amount of donations and nodes in their vicinity fall back to self-investment. A network where both sources lack a donor core is dominated by self-investers. In real world scenarios this can occur when one node is a major hub and the other is located on the periphery. According to our results node on periphery will receive minimal amount of aid. Peripheral location of a node often indicates developing economy, which is sensitive to disturbances like large scale epidemics and hence is reliant on external help. As outlined above global organisation like WHO can take on the role of donor core by contributing initial donations. One exception from this general rule are neighbouring sources. In this set up each source participates in a high fraction of transmission paths from the second infected, thus investment in it diminishes importation risk from both infected. In this setup self-investment is far less prevalent, which means that disease spreading from source to neighbouring nodes does not immediately result in increase of self-investment.

When node's actions are restricted certain recurrent patterns arise. In selfish optimisation mode when detained from investment in source the system always evades to self-investment. In pro-social optimisation mode self-investment is not a beneficial strategy, even when invest-

ment in the source is not an option. Instead neighbours of the source are immunised to act as a protective barrier against disease export. This is analogous to the concept of ring vaccination, which is well known in public health. Nevertheless this strategy requires a much bigger amount of resources to effectively contain the disease as, depending on the network topology, degree of source can be high. In cases where sources have a very low degree their neighbours can receive considerable amount of resources even in a selfish network.

The model at hand is a simplification, which does not include all facets of the real world. While some of those will be difficult to capture in a model, like political relations, skewed risk perception and public pressure, others can be implemented in the future, e.g differing import cost for countries, explicit modelling of the spreading, link cutting, cheating behaviour and different cost functions of notes in the same network (selfish or pro-social). Nonetheless caution needs to be applied, as it is easy to overcomplicate a model. We see simplicity of proposed methodology as a virtue rather than a limitation, as it is reduced to essential features which are necessary to describe the problem with as little assumptions as possible. Some general rules of action can be derived from it, like the lack of difference between pro-social and selfish optimisation under specific circumstances or reduction in utility of self-investment when importation cost are increasing. This emphasises how current regional, national and egocentric thinking conflicts with optimal solution and can lead to suboptimal outcome of pandemic for all participants of the network. We hope that our findings in context of selfish optimisation can be used to persuade governments to rethink their strategic assessments and also give overarching organisations hints on how to manipulate a selfish network to proceed towards a more globally optimal solution.

Chapter 7

Summary and outlook

Pandemics and worldwide outbreaks present a constant threat to society as a whole and to countries situated in high risk regions, e.g. the regions which harbour big populations of mosquitos or where people live in close contact with animals. Thus, it is a heavily researched topic which experienced a boost in research output in the last decade. Insights in this field are highly important for public health, economy and crisis response. Hence, we believe that the results presented in the work at hand are of high relevance for current and future pandemics and can aid decision making in context of the countermeasure deployment. In this thesis we have presented a set of methods to study multiple facets of epidemic outbreaks. We have proposed two context sensitive centrality measures to judge the role of a node in a pandemic situation. Furthermore, we have introduced a method to estimate the threat posed to a specific location by an outbreak. Finally, we established a framework to study resource allocation in the context of pandemic situations and presented multiple general insights in the process of pure cost based decision making. We have published implementation of these algorithms in Julia and Python and encourage all interested parties to use proposed methods.

The methods presented in the work at hand are based upon the concept of effective distance introduced by [17]. Brockmann et. al. propose that the likelihood of an agent to travel across a link can be calculated using the flux over this link. From this probability the effective distance between two neighbouring nodes can be calculated. Effective distances along a path comprise the length of the respective path. Using these, an effective distance based shortest path tree can be constructed. In subsequent paragraphs when using the terms 'shortest paths' and 'shortest path trees' we imply the effective distance as the underlying metric.

As has been pointed out in the previous chapters, conventional centrality measures do not take the specifics of the outbreak into account and thus assign a node the same importance regardless of the outbreak situation. We present two context sensitive metrics, which report the importance of a node, given a specific outbreak location or region. As we have demonstrated in chapter 4, some nodes have a highly variable role in different outbreak scenarios. First metric, *scope*, accounts for the number and size of the nodes located downstream from the node on the effective distance tree. It characterises what fraction of the population of the network is reached by passing through a particular node. Thus, an entry or exit screening of passengers at this node will have protective effect for this fraction of global population. We call nodes which have a very high scope value and thus are present in the most shortest paths, *gates*. Second proposed metric, *confluence*, indicates the branching of the shortest path tree downstream of the node in question. Hence if a node n has low confluence, the shortest path tree remains a single branch after it passes through the node in question. For countermeasure deployment a subsequent node m can be considered, as by shifting the countermeasure

downstream only the population of n is exposed to higher risk. This might be an unavoidable loss when certain factors prevent factual or efficient deployment of countermeasures in node n.

Both metrics can be used to coordinate containment efforts in an informed manner, with the information tailored to the specific outbreak. A further benefit of these metrics is their independence from any disease parameters. Hence, at early stages where these parameters are not yet estimated or subject to strong variations our metrics can be readily used. As it is important to react timely when a pandemic emerges, we see this as one of the key benefits of the proposed metrics. At later stages of the pandemic or in the case of a pathogen surfacing at multiple locations simultaneously, the proposed metrics can be averaged over all affected locations, thus be adjusted to the situation at hand. We have demonstrated that the hubs can be characterised according to their scope profile. Certain airports show high scope and thus play an important role in case of outbreaks in narrow regions, while playing close to no role during outbreaks in the majority of the world. We call these nodes *specialists*. Using scope or confluence averaged over all possible outbreak locations is especially misleading for these nodes, as the mean value misrepresents the importance in every scenario. A second type of nodes are the generalists, which play a considerable role in outbreaks in most regions of the world, but a rarely the dominant gate. Here an averaged value is of less risk. Nonetheless, we strongly encourage to use all available information if possible. Hub nodes can be clustered according to regions, which promote them to gates. This procedure creates well defined groups of nodes which are relevant in case of an outbreak in a broad geographic region. It also highlights relationships between the nodes. Some nodes fall in the cluster of their own geographic region (e.g. Johannesburg O.R. Tambo is associated with the African cluster), others display a strong connection to a different continent (e.g. Madrid Barajas airport, which is assigned to the South-American cluster). Furthermore some historical ties between countries become apparent through effective distance trees and scope values. When considering effective distance tree rooted at Freetown, Sierra Leone, London Heathrow airport gains a very high scope. If the tree is rooted at Conakry, Guinea, Paris Charles de Gaulle takes on a similar role. While geographically close, these countries have a different history. Sierra Leone was tied to Great Britain, while Guinea was colonised by the French. This is reflected in the scope value assigned to the respective airport.

We have demonstrated that our metric can be used for a specific outbreak situation by applying it to theoretical and real-world scenarios. Using the scope, a ranking of nodes can be created, which indicates important targets for countermeasure deployment. These targets vary across presented scenarios, emphasising that there is no one-size-fits-all solution for different pandemics. The ranking can be limited to the target airports from a certain region to inform the officials of the respective region or any other entity with limited range of actions or bound by other limitations. The simplicity and low computational cost of these metrics make them easy to use as has been demonstrated on the world aviation network. Scope and confluence can be applied to other transportation networks or networks accumulating fluxes in the same fashion.

As became apparent in the Ebola crisis in 2014 there is still no consensus about how import risk should be estimated. Elaborate mathematical models present a reliable approach when parametrised successfully. The latter is a challenge as disease parameters, like R_0, are difficult to estimate in field conditions and tend to vary depending on the season, hygiene and other conditions. Even when reliable estimates are available, extensive simulations need knowledge to be implemented and require time and computational resources to complete. We have proposed a metric based on effective distance trees, which allows to calculate the distribution of risk across the network given a specific outbreak origin. Our approach exploits the idea, that the probability of a traveller to traverse a link is proportional to the flux over

the link. Additionally, we expect a traveller to end its journey at a node with the probability proportional to the size and inversely proportional to the scope of the respective node. These processes have an opposing trend. While the probability to travel to a proximate hub is high, exit probability at a hub is very low. Opposite is true for leaf nodes of the effective distance tree: agents rarely travel this far, but if they do they are guaranteed to exit there. Thus, import risk is distributed across the network in a non trivial manner, so that neighbouring hubs are often at lower risk than the second degree neighbours.

Just as the previous metrics, import risk is independent from disease parameters and can be computed as soon as the outbreak origin is known. The local fluctuations of parameters inside a country play no role at the scale at which import risk is estimated. Import risk is not correlated to any conventional centrality measure as has been described in chapter 5. This emphasises that no centrality measure is a good estimate for the threat posed to a node. As described previously, centrality measures do not consider outbreak specific information, that leads to under- or overestimation of risk in particular outbreak scenarios. We emphasise, that import risk is highly sensitive to the outbreak origin, showing that a threat posed to a node varies strongly and no universal statement can be made about the risk without knowing the outbreak situation.

When outbreak location is not bound to a catchment area of the airport, but is distributed across a wide region, the average import risk can be calculated to account for this. Furthermore import risk into a country, as opposed to an airport, can be calculated. While we advice to use all available information if possible, averaged risk can highlight association of certain airports to specific regions. We have introduced risk profiles for airports and countries composed of averaged import risk from each of 22 regions as listed in [58]. Some countries are associated with regions different than their own, e.g. USA is strongly associated with multiple regions in Oceania. Other countries have no clear association with any specific region, e.g. Great Britain, or are threatened exclusively by their own region, e.g. South Africa. The risk profiles show that airports inside one country can display distinct association patterns, especially if the country has a wide geographical stretch.

To demonstrate how import risk can be utilised in an outbreak scenario we have calculated import risk distributions for multiple hypothetical and real-world outbreak scenarios. When considering the worldwide ranking of endangered locations in a specific outbreak scenario an interesting pattern arises. It becomes apparent that most endangered airports lie in regions close to the outbreak, but not necessarily in the same region. It shows that the closest neighbours are not always at highest risk. But we have to point out that the air transport is not the most common type of travel to reach a close location. Thus, when judging risk to geographically proximate locations additional transport networks or approximation models should be considered. Inside each region there exists a set of hubs, which always rank highest. In Europe those are Frankfurt International (FRA), Paris Charles de Gaulle (CDG), London Heathrow (LHR) and Amsterdam Schiphol (AMS). Depending on the outbreak location only a subset of these is observed in the worldwide top ten ranking, but over multiple evaluated scenarios we confirmed that certain regional hubs are consistently at higher threat than their peers. Interestingly, historical ties are also reflected in the distribution of import risk. The import risk ranking for Ebola outbreak shows three European airports among the top ten worldwide. Among them are CDG and Brussels-Zaventem. Both countries, France and Belgium, have historical connection to Guinea, which was one of most severely affected countries.

We must emphasise, that the import risk reflects the probability of a single infected, who has left the country, to end its journey at the respective airport. It does not reflect the absolute risk of the epidemic. If epidemic is not contained and continues for sufficient amount of time eventually each country in the network is expected to have imported cases.

Severity of the outbreak and the disease are likely to influence the number of infected exiting the country. Furthermore, the number of asymptomatic cases can increase the amount of infected travellers. Therefore import risk must be viewed as the distribution of risk across the network rather than the final, absolute probability of case importation.

When the risks are established, actors on the network need to decide how to respond to the epidemic at hand. We propose a simple model to study this question, which reveals general laws observed over a broad spectrum of parameters, susceptibility gain functions and network topologies. We have implemented two modes of optimisation: selfish and pro-social, which reflect the attitude of the agents in the system. Pro-social optimisation can be compared to the resource allocation of a global benevolent entity which aims to reduce the cost inflicted to all nodes on the network. Results discussed below apply to the selfish system if not explicitly stated otherwise. Players in selfish mode minimise only their own cost, thus each node has a different cost function due to a different position on the network. The determining factor for strategic decisions of selfish nodes is the import risk from the outbreak sources. Consequently, depending on the parametrisation, for some nodes it is optimal not to contribute any resources for the disease prevention. In pro-social case all nodes optimise an identical cost function, therefore the conflict inherent to selfish optimisation is not present here. If the outbreak is seeded by a single source, the latter is the only viable target for resource allocation, regardless whether the optimisation is selfish or pro-social. This is confirmed by computational and analytical results alike. The difference between optimisation modes lies in the amount of the resources allocated to the infected. While the susceptibility of the source is reduced to zero in pro-social optimisation, residual risk is tolerated in a selfish system. The precise amount of tolerated residual risk depends on the parameters of the system.

According to presented analytical results, the nodes which present viable targets for investment have two crucial properties. First, they contribute to a considerable subset of paths which account for a high import probability. Second, they have a low susceptibility, thus already exhibit a low probability of disease transmission across the node. Since investment reduces the susceptibility, the process is self-reinforcing. It is evident that a node, which already received any investment, will be favoured for further investment. Consequently, an initial commitment of resources can influence the final equilibrium of the system. The latter is true for linear topologies, like chains and rings, but is not observed on graphs with multiple possible transmission paths. The final equilibrium of such graphs is independent from the initial conditions. On any topology there are at least two nodes which are present in all possible transmission paths: the source of infection and the target node itself. While all nodes share the infected as a viable investment target, it is unlikely that other nodes are a good target for any node but itself. Following from the above, after multiple iterations the source node is likely to have lower susceptibility than any other node in the network, making it the most viable investment target in subsequent steps. When multiple sources are present in the network the situation becomes more complex. Each source is a part of a subset of paths, while the target node itself is present in all possible transmission routes. Nevertheless, source nodes are relevant to other nodes beside the target node, thus likely to receive investment from a higher number of nodes.

While a multi-source system is not analytically solvable with reasonable effort, we can observe its equilibrium in computational results. Multiple strategic regions in the network emerge in this case, with most distinct changes in the selfish optimisation scenario. Similar to [103], we observe a conflict of interest between selfish and pro-social optimum. As outlined above, in pro-social case the position of the nodes plays no role in their investment decision. The conclusion, that the source with lowest susceptibility presents a most favourable target, is confirmed by the final equilibrium. Outbreak sources are immunised in a following order:

first the susceptibility of a source presenting a higher threat to a higher number of nodes is reduced to zero, only thereafter does the second source receive any resources. This is consistent from observations made in different resource allocation models [64, 93].

In the selfish optimisation mode a more complicated picture arises. Nodes in close proximity of an outbreak source and thus at high probability of case importation prefer to invest in the respective source. Nodes which have approximately equal probabilities of case importation from multiple sources prefer to invest in themselves. This is accordant with the analytical insights derived for the general case. For a node which is equidistant to two sources each source contributes to approximately 50% of the import probability. For nodes in close proximity of a source the latter contributes to a higher fraction of import probability, thus making investment in this source a favourable decision. In pro-social optimisation mode the self-investment is never a preferable strategy for non-infected nodes. Our results indicate that in most generic networks with random outbreak source selection the transmission paths from one source rarely traverse any other infected node, thus the investment in any infected has no significant influence on the import probability from other sources. A linear threshold, separating the region of investment into a particular source and the region of non-investment, can be observed in selfish mode simulations. This threshold is dependent only on the import probability form that particular source and not any other infected. Using this as assumption we have calculated the threshold value analytically. The exact value depends on multiple model parameters: the cost of case importation and the probability of an infected to leave the source.

Concluding from our model, some outbreak sources are likely to receive less investment due to their position on the network. Usually a source has a basin of attraction for investment. In this region the import probability from the respective closest source is considerably higher than from any other infected. Thus, it it favourable for nodes in this region to invest in the closest source. In some networks one source lacks the core of this basin, e.g. if the second source is well connected with the rest of the network while the first isn't. In this case nodes in proximity of the first infected turn to self-investment as the preferable strategy. This considerably increases the fraction of self-investment in the network. Under specific circumstances nodes from both core basins can be lacking, in this case the entire network adopts self-investment as its strategy. Nonetheless, when infection sources are direct neighbours, as in case of a disease present in the catchment area of multiple airports of the same country, self-investment is not the dominant strategy in the network. On the contrary, self-investment is highly diminished. We attribute this effect to partial protection from case importation from source i_1 gained by investing in source i_2 and vice versa.

Remaining parameters of the model also have a big influence on the distribution of strategies in the final equilibrium. A disease which is associated with high importation cost leads to a higher fraction of resources being allocated to the infection sources. Similar effect is achieved by an increase in system-wide resources. The latter can be attributed to a higher drop in susceptibility of the source achieved by the nodes inside of its investment basin. While increasing disease cost can only reduce self-investment, increasing system-wide wealth fully eradicates self-investment as a strategy.

When the nodes are restricted from following their preferred strategy multiple alternatives arise. In general, when restricted from investment into the infected the pro-social players are more likely to employ a strategy resembling the ring vaccination. In this case the resources are allocated to the neighbours of the outbreak sources, thus encapsulating the infection inside the source. Selfish players, on the other hand, fall back to self-investment in this case. When both, self-investment and investment into infected, are unavailable, strategies in selfish and pro-social optimisation modes are similar, differing mainly in the amount of allocated resources. Hence, our model shows that the difference between selfish and globally

optimal solution has a considerable overlap in many scenarios.

Results presented in this work generate new insights about the risk and possible countermeasures during a disease outbreak, which presents a threat to global community. This work and proposed metrics contribute to the understanding of different roles the nodes can play in a pandemic. It emphasises that neither the importance of nodes nor the import risk in varying outbreak scenarios can be described by a single number. Our methodology is positioned in the middle ground between the precise, but difficult to parametrise, models and oversimplified general metrics, which disregard all specifics of a particular scenario. Proposed metrics can be applied at early stages of an outbreak and are computationally feasible, since they do not require compute clusters or long time periods to be evaluated. Our metrics help to navigate in a complicated scenario of a pandemic, make more informed decisions and, ultimately, use available resources more efficiently.

By using the scope, transit nodes can be identified reliably and efforts to contain the disease made by multiple downstream countries can be combined at the respective node. In this case, even if the downstream countries do not own considerable amount of resources, an effective response, e.g. state of the art passenger screening, still can be set in place. The confluence can be used to identify and exploit the bottlenecks in the effective distance tree and estimate the loss of protection of the respective population, if the countermeasures are deployed in a more downstream node for any reason. Such reasons can be unstable situation in the upstream country, refusal to accept aid or political sanctions. Whatever the reason and the situation are, these metrics provide a base for more informed and justified decision making.

We are convinced, that import risk is a reliable metric to use during the onset of a pandemic. Import risk accurately identifies the nodes at the highest threat from the outbreak, setting clear priorities. Combined with methods to estimate the number of infected leaving a country the import risk will reflect the risk of importation in a specific time period. However, we emphasise that if an outbreak is not contained and disease ravages for a long time period the import of cases to any country will be imminent. This is an important matter, which is often disregarded when arguing about endemic and emerging pandemics in poor countries, which are unable to cope with the outbreak when left alone. Sophisticated models can complement insights generated by the import risk at later stages, when disease parameters are reliably estimated and disease dynamic is not changing. Otherwise, early modelling attempts can lead to high variation in final prediction and be of limited use. Such projections, when reported prematurely and misrepresented to the public, can undermine public trust and subsequently divert public attention from an urgent crisis.

The import risk can be extended to account for an uncertain outbreak origin or a broad spatial distribution of infected cases. Thus it can be used to evaluate threats of endemic diseases or pandemics with an unequal burden of infected across multiple locations. Multiple examples of such a situation exist, poliomyelitis and Zika outbreaks are the most pressing ones. Probability distributions acquired through import risk calculations can be complemented by additional information, like the habitat of Aedes mosquitos in case of Zika or vaccination status of the population and the vaccine used in the case of polio. The probabilistic nature of import risk makes it easy to incorporate this metric into further models. The most obvious use of the import risk metric lies in informing the governments and global organisations about import probability into countries, states and districts, as opposed to single airports. In this case the import risk can be aggregated to reflect the threat for a broader region, weighing different parts according to users needs. The weighing can be combined with previously mentioned specifics of the local population or the environment. The coarse-grained perspective can also disclose relationships between the geographical regions. We have demonstrated, that certain regions export cases in their direct proximity, while others

pose a threat almost uniformly distributed across the world. Thus, knowledge about the type of export behaviour of the region can help to design a fitting response, well in advance or timely after the outbreak.

Our results concerning the resource allocation demonstrate that the currently dominating perception about resource hoarding and self investment as the safest way through a pandemic, are often false. We have shown that investment into infected is the best strategy in many cases, selfish and pro-social. Thus, investment in the affected region is not an altruistic act, but an efficient strategy of disease containment from a selfish perspective. As most pandemics strike subsequently, the likelihood of a disease emerging in two locations simultaneously is low. Consequently, decision making mostly takes place in the context of a single source outbreak. In such cases investment in the outbreak source is the optimal strategy for any investing country. The importation of disease and an outbreak in the adjacent nodes does not justify self-investment, it in fact makes it less beneficial. However, we have to point out that contrary to the assumption in the model the decision process about catastrophic and humanitarian aid can stretch over long time periods. Thus during the negotiations a new catastrophe can arise and change the situation.

Our work confirms that the ring vaccination is a valid strategy, though it is not the optimal one. Multiple circumstances can make investment in certain countries impossible as has been demonstrated by Nigeria, Afghanistan and Pakistan [2, 44]. Our model indicates that applying the ring vaccination is the second best strategy in this scenario when resources are allocated pro-socially. Selfish optimum deviates from this, which is not relevant for diseases like poliomyelitis, where the response and the resources are managed by the World Health Organisation (WHO), with little participation of non-affected member states. While a high scope of a node makes it a better target than a node with low scope, we do not observe any investment in the high scope targets except for the infected himself. Our analytics suggest, that to make investment in a node beneficial, its scope has to be close to one in its magnitude or the susceptibility of the node must be low prior to the decision making. While investment in high scope nodes can reduce import probability, it can never prevent the importation of the disease if it is not contained, due to less probable existing paths. Consequently, investment in high scope can buy time rather than resolve the problem entirely. Nonetheless, in circumstances that strongly violate the assumptions of our model the investment in high scope nodes can be considered, e.g. the disease is expected to be contained in a fast manner, so that probability reduction is a sufficient protection.

While in the work at hand the resource allocation was studied in context of a disease outbreak, the methodology is sufficiently general to be applied to any spreading phenomena which follows the rules of a contagion process. Many of such phenomena originate from social field, e.g. popularity of brands, spreading of ideas, rumours and social uproar. Thus, the study can be extended to any of these fields and our results apply if the spreading of the process can be contained by means of resource commitment. For example, rumours can be combated by the education and clarifications of facts. When we consider all these phenomena as contagious processes in the model, simultaneous emergence of two outbreaks becomes more likely.

As has been emphasised, the model at hand is not a classical public good game, but it shows features of a public good dilemma. We can observe a disparity between the globally optimal solution and the selfish optimum, where in the latter nodes contribute less to the 'public good' than is required to achieve the global optimum. If multiple outbreaks arise simultaneously, the optima diverge stronger. Emergence of self-investment is the feature of a selfish system, this strategy is never optimal in pro-social context. Furthermore, certain outbreak sources can be at disadvantage compared to other outbreak sources through their position on the network. In these cases they will receive no investment, which is the worst

possible scenario for pro-social optimisation. To counterbalance this, global benevolent actors can contribute resources to the disadvantaged source to reduce its susceptibility. Our model suggests that it will encourage additional investment into the respective outbreak origin. To determine whether a node is at disadvantage the distribution of import probabilities need to be calculated. As demonstrated in our work, the absence of core basin of investment is a clear indicator for investment deficit in the final state.

Our model demonstrates that pure cost based approaches can pose ethical issues even if optimised pro-socially. In case of two outbreak sources and low system-wide resources the model identifies resource commitment to a single source as the optimal strategy, consequently leaving the second outbreak source without aid. This poses an ethical dilemma and the questions whether such a strategy should be implemented, even if it is optimal in terms of cost. This strategy can can provoke opposition not just from the abandoned country alone, but from its neighbours, who are left with higher probability of case importation. Furthermore, our model shows that the distribution of the resources, available for disease containment, among the nodes of the network plays no role for the final amount of recourses received by an outbreak source. The amount of resources in the system as a whole, on the other hand, has a great influence on the outcome. A selfish system with low resources has a strong tendency to self-investment. Such behaviour can drive the system into a weaker economic state, since the economy of a country left without aid during a pandemic will take severe damage. This is a spiralling process, leading to overall decrease in systems wealth and stability.

In summary, we have presented tools to study epidemic outbreaks and prepare countermeasures, prior or at early stage of a pandemic. While we have provided tools to plan outbreak response, we also contributed to a more basic understanding of efficient resource allocation. We encourage all parties to use presented insights to plan and coordinate deployment of pandemic countermeasures. We earnestly hope that this work will help to rebut the local thinking when it comes to global pandemics, eliminate cost inefficient practices and convince global players that aiding infected is a valid option, even from a selfish perspective.

There are multiple possibilities to expand and broaden the research based on the work in this thesis. While we have proposed the metrics, they need to be evaluated on different network topologies and disease models and be compared to existing intervention models. Similar evaluations can be done with the resource allocation patterns reported by our model. In this work, the resource allocation model was kept very simple, reduced to the features indispensable to describe the studied phenomenon. After we have extensively studied the simple model, additional features can be included. Our work was limited to the investigation of the behaviour of network nodes who can act either selfishly or pro-social. To account for the complex interplay between countries, NGOs and other donors, multiple non-connected, altruistic agents can be added to the network. Also, a mixed network of agents with different attitudes can to be investigated to quantify the effect of free-riding by the selfish agents. As has been shown in multiple game theoretic studies, cooperation can arise in repeated games with a punishment system in place. In context of resource allocation the possible punishment includes loss of status, monetary penalties or exclusion from beneficial treaties. It is an important question whether cooperation can be enforced by punishment in context of resource allocation and pandemic countermeasures. Further elements of classical game theory can be included in the model, e.g. cheating behaviour or coordination of players. Players can also exhibit preferences in their willingness to donate resources to a certain affected node, e.g. influenced by political relationships, controversies in cultural and political systems. As resource allocation is rarely detached from other political matters, multi-game view on the question of disease containment can facilitate valuable insights about the interactions of many political and social processes at the same time scale. A big challenge which

should be approached is the implementation of resource allocation and disease spreading on similar time scales. Not only to account for extensive negotiations, but also to consider the delays and logistical challenges connected with physical resource allocation.

Pandemic spreading and prevention remains an active and highly researched topic. Multiple approaches and methodologies coexist and, as has been demonstrated in this thesis, often lead to very similar results. The benefit of new, groundbreaking insights in pandemic research is high, with potential to save millions of lives. We hope that research presented in this thesis will help achieving this ultimate goal.

Chapter 8

Appendix

8.1 Relation between scope and betweenness centrality

From the basic logic, scope and betweenness centrality are very similar. We want to demonstrate that under certain conditions they are equivalent up to a proportionality factor. Note that shortest paths for both metrics need to be calculated based on effective distance. We need to rewrite betweenness centrality in notation used for scope. Classic notation is

$$C_b(n) = \sum_{st} \chi^i_{ts}$$

where $\chi^i_{ts} = 1$ if shortest path from s to t traverses i and 0 otherwise. In notation used for scope betweenness centrality reads

$$C_b(n) = \sum_{s=1}^{N} \sum_{k \in \Theta} 1$$

where Θ is the subset of nodes which is accessible on shortest path tree via node n. Scope is defined as

$$s_n(Y) = \sum_{i \in Y} f_i \sum_{k \in \Theta} \eta_k$$

First assumption is that cumulative flux over each node is equal, $\sum_k F_{nk} = F^*$ for all n. We can further rewrite scope as

$$s_n(Y) = \frac{F^*}{\Phi} \sum_{i \in Y} \sum_{k \in \Theta} 1$$

Hence for the case of equal fluxes over all nodes and $Y = \{0, 1, ..., N\}$ we obtain following proportionality

$$C_b(n) = s_n(Y) \frac{\Phi}{F^*}$$

8.2 Scope based clusters

Following list contains full description of region-based clusters presented in figure 4.7

1. Middle Africa, Western Africa, Eastern Africa
2. Southern Asia
3. Western Asia, Southern Asia

4. Southern Africa
5. Central Asia
6. Northern Europe
7. Melanesia, Australia, New Zealand
8. South-Eastern Asia
9. Northern America
10. Central America
11. Western Europe
12. Northern America, Caribbean
13. Eastern Asia
14. Southern Europe
15. Eastern Europe
16. South America

8.3 Circumstances under which exit probability will not increase along one branch

Exit probability is defined as $q_n(i) = \eta_n / s_n(T_i)$. We consider a case $q_{n-1}(i) > q_n(i)$ which leads to a final formula

$$\frac{\eta - 1}{\eta_n} > \frac{s_{n-1}}{s_n}$$

As per definition $s_{n-1} > s_n$, hence the right side is larger than one. Thus minimal condition for exit probability to increase along a branch is that population of $n-1$ is larger than population of n.

8.4 Taylor expansion of the cost function

As defined in 6 the full cost function is

$$C_n(r_n) = \sum_k R_{kn} + \left(\Gamma p_\infty(n|i) + (1-\Gamma) Q_n \delta_{ni}\right) C_I$$

We are interested in $C_n(r_n^*)$ where $(r_n^*)_k = (r_n)_k + \Delta(r_n)_k$ and $(r_n)_m = (r_n^*)_m$ for all $m \neq k$. Thus we use Taylor expansion to approximate it. Note that $\sum_m (r_n)_m < \rho_n$, this way we ensure that $(r_n)_k$ is independent from $(r_n)_m, m \neq k$. Further we assume linear function of Q of a form $Q_n = 1 - \sum_m R_{nm}$ and $\Delta(r_n)_k = \Delta R_{kn}$. First derivative of the cost function has a form

$$\frac{\partial C_n}{\partial (r_n)_k} = 1 + C_I \left(\Gamma \frac{\partial p_\infty(n|i)}{\partial (r_n)_k} + (1-\Gamma)\delta_{ni} \frac{\partial Q_n}{\partial (r_n)_k} \right)$$

Partial derivative of the import risk differs depending on what node is k. If $k \neq n$

$$\begin{aligned}\frac{\partial p_\infty(n|i)}{\partial (r_n)_k} &= Q_n q_n \sum_{\omega \in \Omega_k} \left(\prod_{m=1}^{L} S_{mm-1} \prod_{m=1, m\neq k}^{L} Q_m \right) * -1 \\ &= -\frac{Q_n q_i}{Q_k} \sum_{\omega \in \Omega_k} \left(\prod_{m=1}^{L} S_{mm-1} \prod_{m=1}^{L} Q_m \right)\end{aligned}$$

where Ω_k is a set of paths containing traversing node k and $\omega = \{n_0, n_1, ...n_L\}$ is a path of length L from infected node $i = n_0$ to the acting node $n = n_L$ and $S_{mm-1} = P_{mm-1}(1 - q_{m-1})$. Second part of the equation differs depending on whether $k = n$ or not. If $n \neq k$, $\partial Q_n / \partial (r_n)_i = 0$, otherwise it is 1. Further if $k = n \neq i, \delta_{ni} = 0$. In this case solution is equivalent to $k \neq n$. All higher derivatives are equal to 0. As discussed in the main manuscript strategic solution for the source node is trivial, hence we will not consider this case using Taylor expansion. Solution of the Taylor expansion for $k \neq i$ is as follows:

$$\begin{aligned}C(r_n^*) &= \sum_m R_{mn} + \Delta(r_n)_k + \left[\Gamma \left(p_\infty(n|i) - \frac{Q_n q_n}{Q_k} \sum_{\omega \in \Omega_k} \prod_{n=1}^{L} \hat{S}_{nn-1} \right) + (1 - \Gamma) Q_n \delta_{ni} \right] C_I \Delta(r_n)_k \\ &= C(r_n) + \Delta(r_n)_k - C_I \Gamma \frac{\sum_{\omega \in \Omega} p(\omega)}{Q_k} \Delta(r_n)_k\end{aligned}$$

where $\hat{S}_{nm} = (1 - q_m) P_{nm} Q_m$. Note that $Q_n(R)$ is discontinuous for $\sum_m R_{nm} > 1$. Further $R_{nm} \geq 0$ for all m, n so that above formulas are defined in range $\sum_m R_{nm} \in [0, 1]$.

8.5 Comparison of strategic decisions

8.5.1 Invest in n

Assume node n has two possible strategies r_n and r_n^* with $(r_n^*)_k = (r_n)_k + \Delta(r_n)_k$ and $(r_n^*)_m = (r_n)_m$ for $m \neq k$. This means that strategy r_n^* allocates more resources to node k than strategy r_n. Further $C_n(r_n^*) < C_n(r_n)$, meaning that strategy r_n^* is preferable one. Using Taylor expansion above we can derive conditions for the above to be true.

$$\begin{aligned}C(r_n) &> C(r_n) + \Delta(r_n)_k - C_I \Gamma \frac{\sum_{\omega \in \Omega} p(\omega)}{Q_k} \Delta(r_n)_k \\ 1 &< C_I \Gamma \frac{\sum_{\omega \in \Omega} p(\omega)}{Q_k} \\ \frac{Q_k}{C_I \Gamma} &< \sum_{\omega \in \Omega} p(\omega)\end{aligned}$$

Constrains imposed by Taylor expansion and model definition remain.

8.5.2 Invest in a or invest in b

Assume node n has a strategy r_n, from which it considers to deviate. There are two alternative strategies r_n' and r_n^* which differ from the default by $\Delta(r_n)_a$. In strategy r_n' more investment is directed to node a, thus $(r_n')_a = (r_n)_a + \Delta(r_n)_a$ and $(r_n')_m = (r_n)_m$ for $m \neq a$. In strategy r_n^*

additional investment is donated to node b, thus $(r_n^*)_b = (r_n)_b + \Delta(r_n)_b$ and $(r_n^*)_m = (r_n)_m$ for $m \neq b$. Further $C_n(r'_n) > C_n(r_n^*)$, meaning that strategy r_n^* is the preferable strategy. Using Taylor expansion above we can derive conditions for the above to be true.

$$C(r'_n) > C(r_n^*)$$
$$C(r_n) + \Delta(r_n)_a - C_I\Gamma\frac{\sum_{\omega\in\Omega(a)} p(\omega)}{Q_a}\Delta(r_n)_a > C(r_n) + \Delta(r_n)_a - C_I\Gamma\frac{\sum_{\omega\in\Omega(b)} p(\omega)}{Q_b}\Delta(r_n)_a$$
$$\frac{\sum_{\omega\in\Omega(a)} p(\omega)}{Q_a} < \frac{\sum_{\omega\in\Omega(b)} p(\omega)}{Q_b}$$

Constrains imposed by Taylor expansion and model definition remain.

8.6 Probability to receive investment

8.6.1 Probabilities on a chain

Assume we have a chain of length L with infected node at index 0. At time step $t = 0$ all nodes have equal susceptibilities. In case of a chain there is only one path on which disease can spread, hence for every node n_a at index a all nodes n_k with $k \leq a$ are valid targets for investment. Note that we have proven using restriction that a node can only change its investment in one node at a time. Nonetheless on a chain same is likely to be true if this restriction is lifted. At each time step a node, which is given the opportunity to update its strategy, is chosen with uniform probability over all nodes on the chain. Probability of an acting node n_a to invest in any node n_m with $m \leq a$ is $1/a$ and 0 for all nodes $m > a$. Hence probability for each node n_k to receive investment at time step $t = 0$ is

$$\pi(k, t = 0) = \frac{1}{L}\left(\frac{1}{k} + \frac{1}{k+1} + \frac{1}{k+2} + ... + \frac{1}{L}\right) = \frac{1}{L}\left(H_L - H_{k-1}\right)$$

where H_L and H_{k-1} are harmonic series. To calculate probability that n_k receives investment at $t = 1$ two additional scenarios must be considered.

- At $t = 0$ node n_m with $m > n$ receives investment

$$\frac{1}{L}\sum_{m=k+1}^{L}\left(H_L - H_{m-1}\right)$$

- At $t = 0$ node n_m with $m < n$ receives investment

$$\frac{1}{L}\sum_{m=1}^{k-1}\left(H_L - H_{k-1}\right)$$

Note that due to posed restriction investment can be granted to only one node per step. Depending on which scenario occurred at $t = 0$ subsequent actions will differ.

- If node n_k received investment at $t = 0$ it is guaranteed to receive investment at $t = 1$ if a node n_a with $a \geq k$ is chosen as acting node. If $a < k$ no investment in k will be

performed

$$\frac{L-k+1}{L}$$

- If at $t = 0$ node n_m with $m > k$ received investment node n_k has the probability to receive resources from node n_a with $k \leq a < m$ if one of those is chosen to be the acting node

$$\frac{m-k+1}{L}\left(H_{m-1} - H_{k-1}\right)$$

- If at time $t = 0$ a node n_m with $m < k$ was granted resources k will not receive any investment in subsequent rounds

We can combine all of the above to calculate probability of n_k receiving investment at $t = 1$

$$\begin{aligned}\pi(k, t=1) =& \frac{1}{L}\left(H_L - H_{k-1}\right)\frac{L-k+1}{L} \\ &+ \frac{1}{L}\sum_{m=k+1}^{L}\left(\left(H_L - H_{m-1}\right)\frac{m-k+1}{L}\left(H_{m-1} - H_{k-1}\right)\right) \\ =& \frac{L-n+1}{L^2}\left(H_L - H_{k-1}\right) + \frac{1}{L^2}\sum_{m=k}^{L-1}\left((m-k+2)(H_L - H_m)(H_m - H_{k-1})\right)\end{aligned}$$

8.6.2 Probabilities on a more general topology

Assume a general network consisting of outbreak source i, two components $\mathcal{N}_1$ and $\mathcal{N}_2$ and a node j which connects components $\mathcal{N}_1$ and $\mathcal{N}_2$ as depicted in fig. 8.1. Hence node i is part of all transmission relevant paths, while j is part of all transmission paths leading to nodes in compartment $\mathcal{N}_2$. As in previous considerations when investment game is initiated all nodes share the same susceptibility. At time step $t = 0$ one node n_a is chosen to update its strategy. Following cases can arise when a node n_a is chosen to be acting node

- if node $n_a \in \mathcal{N}_1$ it is a viable strategy to invest in node i and himself with equal probability
- if node $n_a \in \mathcal{N}_2$ it is a viable strategy to invest in node i, node j and himself with equal probability
- if $n_a = i$ only viable option is to invest in self
- if $n_a = j$ viable options are node i and node j with equal probability

Following the cases above we can calculate the probability of nodes i and j to receive investment

$$\pi(i, t=0) = \frac{1}{N}\left(0.5|\mathcal{N}_1| + 0.3|\mathcal{N}_2| + 1 + 0.5\right)$$
$$\pi(j, t=0) = \frac{1}{N}\left(0.3|\mathcal{N}_2| + 0.5\right)$$

For subsequent step it matters whether i or j received investment at $t = 0$.

- If at $t = 0$ i received investment it will be always granted resources in subsequent steps until its susceptibility drops low enough

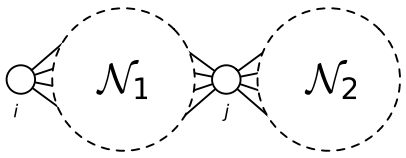

Figure 8.1: General network topology used to calculate probability to receive investment for nodes i and j. $\mathcal{N}_1$ and $\mathcal{N}_2$ are network components containing an arbitrary number of nodes. Node i is outbreak source, node j connects network components $\mathcal{N}_1$ and $\mathcal{N}_2$. Being the source node i is present in all relevant transmission paths. Node j is present in all transmission paths for nodes in component $\mathcal{N}_2$.

- If at $t = 0$ j received investment every acting node $n_a \in \mathcal{N}_2$ chosen at $t = 1$ will donate its resources to j
- If any other nodes received investment and is drawn as acting node at $t = 1$ it will donate further resources, if any, to itself

Following from the above we can derive probability of i and j to receive investment at time step $t = 1$

$$\begin{aligned}\pi(i,t=1) =& \pi(i,t=0) + (1-\pi(i,t=0))\frac{1}{N}\left[\pi(j,t=0)\,(0.5|\mathcal{N}_1|+1) + \right.\\ & 0.5\frac{|\mathcal{N}_1|}{N}\,(0.5(|\mathcal{N}_1|-1)+1.5+0.3|\mathcal{N}_2|) + \\ & \left. 0.3\frac{|\mathcal{N}_2|}{N}\,(0.5|\mathcal{N}_1|+1.5+0.3(|\mathcal{N}_2|-1))\right]\\ \pi(j,t=1) =& \pi(j,t=1)\frac{1}{N}\,(|\mathcal{N}_2|+1) + \\ & (1-\pi(j,t=0)-\pi(i,t=0))\frac{1}{N}\,[0.3(|\mathcal{N}_2|-1)+0.5]\end{aligned}$$

8.7 Observed investment threshold

Using first derivative of the cost function we can calculate the extreme points of the function with respect to one parameter. We are particularly interested in minima with respect to investment into a specific node. These are the points where investment turns from beneficial into disadvantageous.

$$\frac{\partial C(r_n)}{\partial (r_n)_k} = \Delta(r_n)_k - \frac{C_I\Gamma}{Q_k}\sum_{\omega\in\Omega(k)} p(\omega)$$

$$\frac{1}{C_I\Gamma} = \frac{\sum_{\omega\in\Omega(k)} p(\omega)}{Q_k}$$

As already stated above following assumptions were made:

- linear susceptibility function of a form $Q_k = 1 - \sum_n Q_{kn}$ which is defined between inside the interval $[0,1]$
- investment into single nodes are independent, thus $\sum_k (r_n)_k \ll \rho_n$
- $k \in \mathcal{I} \vee k \neq n$ where $\mathcal{I}$ is the set of infected nodes

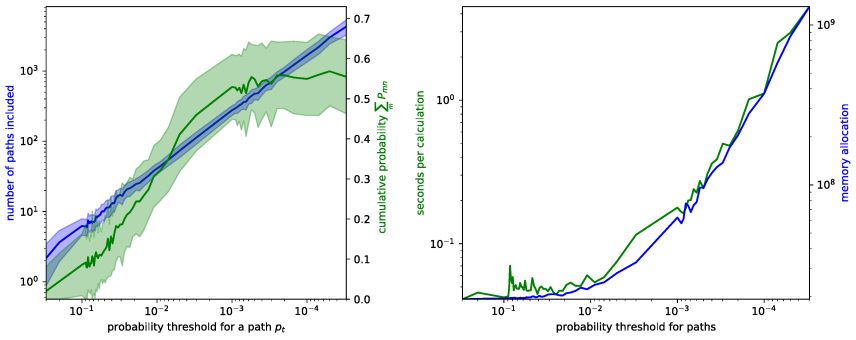

Figure 8.2: Screening of path probability threshold. **A** shows overall number of paths included (blue) and the overall import probability accounted for by the paths included in the calculation (green). Shaded area represents one standard deviation. Each threshold value was evaluated on 100 randomly generated Erdős–Rényi networks, source of the outbreak was chosen at random, each network contained 100 nodes.Care was taken to ensure that no nodes were disconnected from the network. As has been shown in 5.2 in Erdős–Rényi long paths have to be considered to account for high fraction of import risk.While number of paths increases steadily accounted probability saturates at approximately 0.6. **B** shows the increase in computational time and memory allocation. Both is increasing nearly exponentially.

8.8 Choice of path threshold and filtering algorithm

For computational purposes we have to limit the number of paths considered during the calculation of import probability from source to the acting node. First we disregard all paths containing loops. Second we introduce a threshold below which a path is not entering the calculation. To incorporate both restrictions we use a modified version of Yen algorithm. Yen algorithm relies on Dijkstra algorithm, which was develop to handle networks in which link weights are additive. In our model path lengths are judged according to the probability of an agent traversing those. Further exit probability is entering the calculation. Hence link weights are multiplicative rather than additive. Our modified algorithm uses the probability to proceed along a link S_{ij} and translates it to a link weight $w_{ij} = -log(P_{ij})$. It then follows a classical Yen algorithm as described in **[?]**. Instead of finding k shortest paths, our modified algorithm proceeds until first path below a threshold is reached. Threshold is defined as $T = -log(p_t)$ where p_t is a threshold in terms of probability. As soon as such a path is found it is discarded, all paths above the threshold for current node are found. Hence algorithm proceeds to the calculation of paths for the next node.

To decide on appropriate threshold we scanned the threshold monitoring the quality of the approximation and its computational costs. Note that analytically $\sum_m (P^t)_{nm} = 1$ regardless of t. Hence if all paths are regarded we expect the probability to reach 1. As can be seen in fig. 8.2 when paths are thresholded cumulative probability saturates at about 0.6, meaning that the residual is constituted from a high number of paths with low probability. Computational time increases nearly exponentially, hence small increase in probability coverage comes at a high prise. To reduce computational time but harness highest possible probability coverage we set the threshold $p_t = 4 * 10^{-4}$.

8.8.1 Modified evolutionary strategy algorithm

With increasing network size the strategy space for nodes grows, making it unfeasible to calculate the exact optimal solution. Therefore a heuristical approach was used to optimise the cost function. We used a modified version of an evolutionary strategy algorithm [16]. Base algorithm generates strategy vectors randomly, evaluates costs and chooses the one with the lowest. In a modified version a certain fraction of best scoring strategies is kept until the next round. In subsequent round fraction of new strategies is produced by shuffling multiple best performing strategies from previous round. Additionally some strategies are generated randomly the same way as in original algorithm. Cost are evaluated for all generated strategies and strategy with lowest cost is adopted.

Modified version of the algorithm operates with three sets of strategies: Θ, which contains randomly generated strategies, Φ, which contains best performing strategies from previous round, and Ψ, which is a set filled with strategies based on strategies from Φ. Assume a node is reconsidering its strategy. To do this a number of strategies N is generated. In the first round ($t = 0$) a set of strategies Θ is generated randomly where $|\Theta| = N$. Cost for each strategy are evaluated and the strategies are ranked accordingly. Best strategy is adopted by the node. A number of top performing strategies are kept as set Φ with $|\Phi| = M$. At the second round $t = 1$ set Ψ is filled with strategies generated using strategies from Φ. To do multiple strategies are chosen from set Φ as template strategies, e.g. r,r^* and r'. A new strategy $r^{\#}$ is produced by drawing vector value at each index n from values with same index in one of the template strategies, i.e. $(r^{\#})_n \sim U(\{r_n, r^*_n, r'_n\})$ for all n. This procedure is repeated to fill a set Ψ so that $|\Psi| = L$. Additional strategies are generated by random to fill the set Θ so that $\Theta = N - M - L$. Finally cost for all N strategies are evaluated, best performing strategy is adopted by the node and all sets are cleared. Set Φ is filled with M best performing strategies from the current round. The algorithm proceeds to the next node.

To ensure that a newly generated strategy complies to the requirement $\sum_k (r_n)_k = \rho_n$ following procedure is employed. Strategy vector r_n is initiated as a zero vector and filled sequentially in random order. For each index i in the strategy vector a number is drawn from uniform distribution so that $x \sim U([0, \rho_n])$. If $\sum_k (r_n)_k + x \leq \rho_n$ x is added to the vector at index i, otherwise $(r_n)_i = 0$. Vector is filled in random order to avoid disproportionally low values at high indices. This procedure is repeated until the vector is filled.

8.9 Tables

8.9.1 IATA code of 100 biggest airports

IATA code	airport name	IATA code	airport name
AMS	Amsterdam Airport Schiphol	ARN	Stockholm Arlanda Apt.
ATL	Atlanta Hartsfield–Jackson Intl. Apt.	AUH	Abu Dhabi International Apt.
BCN	Barcelona–El Prat Apt.	BKK	Bangkok Suvarnabhumi International Apt.

IATA code	airport name	IATA code	airport name
BNE	Brisbane International Apt.	BOG	Bogotá El Dorado International Apt.
BOM	Mumbai Chhatrapati Shivaji Maharaj Intl. Apt.	BOS	Boston Logan International Apt.
BRU	Brussels International Apt.	BSB	Brasília International Apt.
BWI	Baltimore–Washington International Apt.	CAN	Guangzhou Baiyun International Apt.
CDG	Paris Charles de Gaulle Apt.	CGH	Sao Paulo–Congonhas Apt.
CGK	Jakarta Soekarno–Hatta International Airport	CKG	Chongqing Jiangbei International Apt.
CLT	Charlotte Douglas International Apt.	CPH	Copenhagen Kastrup Apt.
CTS	Sapporo New Chitose Apt.	CTU	Chengdu Shuangliu International Apt.
DCA	Washington Ronald Reagan National Apt.	DEL	Delhi Indira Gandhi International Apt.
DEN	Denver International Apt.	DFW	Dallas/Fort Worth International Apt.
DME	Moscow Domodedovo Apt.	DOH	Doha Hamad International Apt.
DTW	Detroit Wayne County Apt.	DUB	Dublin International Apt.
DUS	Düsseldorf International Apt.	DXB	Dubai International Apt.
EWR	Newark Liberty International Apt.	FCO	Rome Fiumicino Apt.
FLL	Fort Lauderdale–Hollywood Intl Apt	FRA	Frankfurt International Apt.
FUK	Fukuoka International Apt.	GIG	Rio de Janeiro–Antonio Carlos Jobim Intl. Apt.
GMP	Seoul Incheon International Apt.	GRU	Sao Paulo–Guarulhos International Apt.
HGH	Hangzhou Xiaoshan International Apt.	HKG	Hong Kong International Apt.

IATA code	airport name	IATA code	airport name
HND	Tokyo Haneda International Apt.	HNL	Honolulu Daniel K. Inouye International Apt.
IAD	Washington Dulles International Apt.	IAH	Houston George Bush Intercontinental Apt.
ICN	Seoul Incheon International Apt.	IST	Istanbul Atatürk Apt.
JED	Jeddah King Abdulaziz International Apt.	JFK	New York J. F. Kennedy International Apt.
JNB	Johannesburg O.R. Tambo International Apt.	KIX	Osaka Kansai International Apt.
KMG	Kunming Wujiaba International Apt.	KUL	Kuala Lumpur International Apt.
LAS	Las Vegas McCarran International Apt.	LAX	Los Angeles International Apt.
LGA	New York LaGuardia Apt.	LGW	London Gatwick Apt.
LHR	London Heathrow Apt.	MAD	Madrid Adolfo Suarez–Barajas Apt.
MAN	Manchester International Apt.	MCO	Orlando International Apt.
MDW	Chicago Midway International Apt.	MEL	Melbourne International Apt.
MEX	Mexico City International Apt.	MIA	Miami International Apt.
MNL	Manila Ninoy Aquino International Apt.	MSP	Minneapolis International Apt.
MUC	Munich International Apt.	MXP	Milan–Malpensa Apt.
NRT	Tokyo Narita International Apt.	ORD	Chicago O'Hare International Apt.
ORY	Paris Orly Apt.	OSL	Oslo Gardermoen Apt.
PEK	Beijing Capital International Apt.	PHL	Philadelphia International Apt.
PHX	Phoenix Sky Harbor International Apt.	PMI	Palma De Mallorca Apt.

IATA code	airport name	IATA code	airport name
PVG	Shanghai Pudong International Apt.	RUH	Riyadh King Khalid International Apt.
SAN	San Diego International Apt.	SAW	Istanbul Sabiha Gökçen International Apt.
SEA	Seattle–Tacoma International Apt	SFO	San Francisco International Apt.
SGN	Ho Chi Minh City Tan Son Nhat Intl. Apt.	SHA	Shanghai Hongqiao International Apt.
SIN	Singapore Changi Apt.	SLC	Salt Lake City International Apt.
SUB	Surabaya Juanda International Apt.	SVO	Moscow Sheremetyevo International Apt.
SYD	Sydney Kingsford Smith Apt.	SZX	Shenzhen Bao'an International Apt.
TPE	Taipei Taiwan Taoyuan International Apt.	TXL	Berlin Tegel Apt.
VIE	Vienna International Apt.	XIY	Xi'an Xianyang International Apt.
XMN	Xiamen Gaoqi International Apt.	YVR	Vancouver International Apt.
YYZ	Toronto Pearson International Apt.	ZRH	Zurich Airport

Table 8.1: 100 biggest airports according to flux

8.9.2 List of generalist and specialist airports among 100 biggest on the WAN

Generalists	Specialists
Abu Dhabi International Apt.	Baltimore–Washington International Apt.
Amsterdam Airport Schiphol	Beijing Capital International Apt.
Atlanta Hartsfield–Jackson International Apt.	Bogotá El Dorado International Apt.

Generalists	Specialists
Bangkok Suvarnabhumi International Apt.	Brasília International Apt.
Barcelona–El Prat Apt.	Brisbane International Apt.
Berlin Tegel Apt.	Chongqing Jiangbei International Apt.
Boston Logan International Apt.	Copenhagen Kastrup Apt.
Brussels International Apt.	Delhi Indira Gandhi International Apt.
Charlotte Douglas International Apt.	Denver International Apt.
Chengdu Shuangliu International Apt.	Fort Lauderdale–Hollywood Intl Apt
Chicago Midway International Apt.	Hangzhou Xiaoshan International Apt.
Chicago O'Hare International Apt.	Ho Chi Minh City Tan Son Nhat Intl. Apt.
Dallas/Fort Worth International Apt.	Jakarta Soekarno–Hatta International Airport
Detroit Wayne County Apt.	Johannesburg O.R. Tambo International Apt.
Doha Hamad International Apt.	Kuala Lumpur International Apt.
Dubai International Apt.	Toronto Pearson International Apt.
Dublin International Apt.	Melbourne International Apt.
Düsseldorf International Apt.	Mexico City International Apt.
Frankfurt International Apt.	Moscow Sheremetyevo International Apt.
Fukuoka International Apt.	Mumbai Chhatrapati Shivaji Maharaj Intl. Apt.
Guangzhou Baiyun International Apt.	New York J. F. Kennedy International Apt.
Hong Kong International Apt.	New York LaGuardia Apt.
Honolulu Daniel K. Inouye International Apt.	Oslo Gardermoen Apt.
Houston George Bush Intercontinental Apt.	Rio de Janeiro–Antonio Carlos Jobim Intl. Apt.
Istanbul Atatürk Apt.	Riyadh King Khalid International Apt.
Istanbul Sabiha Gökçen International Apt.	Rome Fiumicino Apt.
Jeddah King Abdulaziz International Apt.	San Diego International Apt.
Kunming Wujiaba International Apt.	Sao Paulo–Congonhas Apt.

Generalists	Specialists
Las Vegas McCarran International Apt.	Sao Paulo–Guarulhos International Apt.
London Gatwick Apt.	Seoul Gimpo International Apt.
London Heathrow Apt.	Shanghai Hongqiao International Apt.
Los Angeles International Apt.	Shenzhen Bao'an International Apt.
Madrid Adolfo Suarez–Barajas Apt.	Stockholm Arlanda Apt.
Manchester International Apt.	Surabaya Juanda International Apt.
Manila Ninoy Aquino International Apt.	Tokyo Haneda International Apt.
Miami International Apt.	Tokyo Narita International Apt.
Milan–Malpensa Apt.	Washington Ronald Reagan National Apt.
Minneapolis–Saint Paul International Apt.	Xi'an Xianyang International Apt.
Moscow Domodedovo Apt.	
Munich International Apt.	
Newark Liberty International Apt.	
Orlando International Apt.	
Osaka Kansai International Apt.	
Palma De Mallorca Apt.	
Paris Charles de Gaulle Apt.	
Paris Orly Apt.	
Philadelphia International Apt.	
Phoenix Sky Harbor International Apt.	
Salt Lake City International Apt.	
San Francisco International Apt.	
Sapporo New Chitose Apt.	
Seattle–Tacoma International Apt	
Seoul Incheon International Apt.	
Shanghai Pudong International Apt.	

Generalists	Specialists
Singapore Changi Apt.	
Sydney Kingsford Smith Apt.	
Taipei Taiwan Taoyuan International Apt.	
Vancouver International Apt.	
Vienna International Apt.	
Washington Dulles International Apt.	
Xiamen Gaoqi International Apt.	
Zurich Airport	

Table 8.2: Clusters according to airports profiles

8.9.3 Default parameter set

Table below describes parameters used in the model. Listed value is the default used when nothing further is specified. Name used in Julia implementation is included in the last row.

Parameter	default value	description
Network size	100	Number of nodes in the network
Average degree	4	Average number of neighbors in the network
Susceptibility function	linear	function which converts donated resources into susceptibility reduction
Update order	random	order in which nodes are chosen to update their strategy at each step
Duration of strategy deployment	10	to prevent oscillations and allow system to change steadily strategy changes were adopted stepwise; during the adaptation strategies were evaluated and reconsidered as described above
Gamma	0.5	fraction of infected leaving the node
Path threshold	0.0004	Threshold below which paths were disregarded
Maximal rounds	200	Maximal number of rounds performed during the simulation

Parameter	default value	description
Number of generated strategies	100	Number of strategies N generated by the evolutionary strategy algorithm for each node
Number of best strategies kept	30	Number of strategies M kept to be used in subsequent step
Number of shuffled strategies	15	Number of strategies L which are generated by shuffling best strategies of previous round
Number of template strategies	3	Number of strategies to be used as template for creation of shuffled strategies
Susceptibility function exponent	1	Exponent of the susceptibility function; only considered when exponential susceptibility function is used
Cost of infection import	50	Cost connected to one imported case
Outbreak severity	20	Scaling factor to represent the severity of the outbreak
Amount of resources	Single source: $1.4/N$ Double source: $2.4/N$ Tripple source: $3.4/N$	Amount of resources available to the system, N is the number of nodes in the network; in default state all nodes had equal amount of resources
Number of rounds without change before termination	20	Required number of rounds without strategy changes before system is considered to have reached it final steady state

Table 8.3: Default parameters used in the resource allocation model

Chapter 9

Acknowledgements

The work was made possible by a variety of people. I thank my supervisor, Dirk Brockmann, for his scientific guidance and support. I appreciate the freedom I experienced while working on the thesis (the topic, the timeline, the methods), the opportunities to travel to conferences and research partners. The atmosphere he facilitates in the group is something I will surely miss once my time in the Brockmann lab is over.

Further, let me express gratitude to Benni for scientific and less scientific discussions. His cheerful mindset and fascination for science are truly remarkable. He granted me valuable insights into the mind of a physicist, a species I knew little about when starting my PhD journey.

The four years would have been much less enjoyable if not for the current and former members of the group. Thanks to Chen Li for his help during the first years of my thesis and to Frank for his advice in the last ones. Thanks to Julien for answering the weirdest computer related questions I came up with and to Fakhteh for discussions about the analytic approach to my project. Thanks to the entire group for being an awesome, openminded bunch! I owe special gratitude to Clara, Paola, Frank and Benni for proofreading parts of my work.

There are multiple scientists who were not involved in my work directly, but were an inspiration throughout the years. Andreas Voigt was the one to encourage my tendency to ask questions, always having the time and the books I needed. Peter Hammerstein and January Weiner turned an alleged easy-A class into the most memorable lecture of my masters, showing that biology with numbers is much more compelling than without. Another scientist who deserves a separate mention is Alexandra Elbakayan. I hope one day I can inspire young scientists the same way she does.

There are countless things I have to thank my family for, but I will limit myself to the topic. I thank my mother for supporting my every ambition and my creativity, but warmly guiding my wildest ideas through logic. I thank my brother for being a person I could look up to, sometimes yell at, but always talk to.

Finally, I want to thank Micha for more than words could ever communicate. I am grateful for all the times you lifted me up, calmed me down, for the time we spent talking about private and work-related topics. Frankly, for all the time we spend together.

References

[1] The new clipper ship thermopylae. *The Angus (Melbourne, Vic. : 1848 - 1957)*, Wed 13 Jan 1869:6, January 1869. 3

[2] Seye Abimbola, Asmat Ullah Malik, and Ghulam Farooq Mansoor. The final push for polio eradication: Addressing the challenge of violence in afghanistan, pakistan, and nigeria. *PLOS Medicine*, 10(10):1–4, 10 2013. 107

[3] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, January 2002. 14

[4] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *nature*, 406(6794):378–382, 2000. 14, 18

[5] Simon P Anderson, Jacob K Goeree, and Charles A Holt. A theoretical analysis of altruism and decision error in public goods games. *Journal of Public Economics*, 70(2):297–323, 1998. 19

[6] Mathieu Andraud, Niel Hens, Christiaan Marais, and Philippe Beutels. Dynamic Epidemiological Models for Dengue Transmission: A Systematic Review of Structural Approaches. *PLOS ONE*, 7(11):e49085, November 2012. 3

[7] James Andreoni. Why free ride?: Strategies and learning in public goods experiments. *Journal of public Economics*, 37(3):291–304, 1988. 19

[8] FIFTY-SIXTH WORLD HEALTH ASSEMBLY, editor. *Severe acute respiratory syndrome (SARS)*, number 14.16. WHO, WHO, May 2003. 1

[9] Rebecca F. Baggaley, Neil M. Ferguson, and Geoff P. Garnett. The epidemiological impact of antiretroviral use predicted by mathematical models: a review. *Emerging Themes in Epidemiology*, 2:9, September 2005. 72

[10] The World Bank. Population, total | Data, 2016. 28, 59

[11] Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, October 1999. 14

[12] Chris T. Bauch and David J. D. Earn. Vaccination and the theory of games. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36):13391–13394, September 2004. 5, 8, 19

[13] David M. Bell. Public Health Interventions and SARS Spread, 2003. *Emerging Infectious Diseases*, 10(11):1900–1906, November 2004. 41

[14] Daniel Bernoulli and Sally Blower. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. *Reviews in medical virology*, 14(5):275–288, 2004. 16

[15] Hans-Georg Beyer. Toward a theory of evolution strategies: On the benefits of sex - the $(\mu/\mu, \lambda)$ theory. *Evolutionary Computation*, 3(1):81–111, 1995. 21, 22

[16] Hans-Georg Beyer and Hans-Paul Schwefel. Evolution strategies–a comprehensive introduction. *Natural computing*, 1(1):3–52, 2002. 20, 118

[17] Dirk Brockmann and Dirk Helbing. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science*, 342(6164):1337–1342, December 2013. iv, vi, 2, 4, 6, 9, 18, 27, 30, 31, 34, 35, 52, 101

[18] Wouter Van den Broeck, Corrado Gioannini, Bruno Gonçalves, Marco Quaggiotto, Vittoria Colizza, and Alessandro Vespignani. The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infectious Diseases*, 11:37, February 2011. 3, 18

[19] Duncan S. Callaway, M. E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.*, 85:5468–5471, Dec 2000. 14

[20] Hanshuang Chen, Guofeng Li, Haifeng Zhang, and Zhonghuai Hou. Optimal Allocation of Resources for Suppressing Epidemic Spreading on Networks. *Physical Review E*, 96(1), July 2017. arXiv: 1702.08444. 24, 72

[21] S Clarke, A Krikorian, and J Rausen. Computing the n best loopless paths in a network. *Journal of the Society for Industrial and Applied Mathematics*, 11(4):1096–1102, 1963. 16

[22] Brian J. Coburn, Bradley G. Wagner, and Sally Blower. Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1n1). *BMC Medicine*, 7:30, June 2009. 3

[23] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85:4626–4628, Nov 2000. 14

[24] Reuven Cohen, Shlomo Havlin, and Daniel ben Avraham. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.*, 91:247901, Dec 2003. 18

[25] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2015–2020, February 2006. 3, 18, 27, 52

[26] Daniel M Cornforth, Timothy C Reluga, Eunha Shim, Chris T Bauch, Alison P Galvani, and Lauren Ancel Meyers. Erratic flu vaccination emerges from short-sighted behavior in contact networks. *PLoS Computational Biology*, 7(1):e1001062, 2011. 5, 19

[27] Benjamin J. Cowling, Lincoln LH Lau, Peng Wu, Helen WC Wong, Vicky J. Fang, Steven Riley, and Hiroshi Nishiura. Entry screening to delay local transmission of 2009 pandemic influenza A (H1n1). *BMC Infectious Diseases*, 10:82, March 2010. 41

[28] Guilherme Ferraz de Arruda, André Luiz Barbieri, Pablo Martín Rodríguez, Francisco A. Rodrigues, Yamir Moreno, and Luciano da Fontoura Costa. Role of centrality for the identification of influential spreaders in complex networks. *Physical Review E*, 90(3):032812, September 2014. 4, 13, 18, 34

[29] Anthony Dekker. Conceptual distance in social network analysis. *Journal of Social Structure (JOSS)*, 6, 2005. 13

[30] Ernst Fehr and Simon Gächter. Cooperation and punishment in public goods experiments. *The American economic review*, 90(4):980–994, 2000. 6, 19, 72

[31] Paul EM Fine. Herd immunity: history, theory, practice. *Epidemiologic reviews*, 15(2):265–302, 1993. 17

[32] Feng Fu, Daniel I Rosenbloom, Long Wang, and Martin A Nowak. Imitation dynamics of vaccination behaviour on social networks. *Proceedings of the Royal Society of London B: Biological Sciences*, 278(1702):42–49, 2011. 5

[33] Sebastian Funk, Marcel Salathé, and Vincent A. A. Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of The Royal Society Interface*, page rsif20100142, May 2010. 5

[34] Marcelo F. C. Gomes, Ana Pastore y Piontti, Luca Rossi, Dennis Chao, Ira Longini, M. Elizabeth Halloran, and Alessandro Vespignani. Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak. *PLoS Currents*, 6, September 2014. 3, 18

[35] Lawrence O. Gostin, Oyewale Tomori, Suwit Wibulpolprasert, Ashish K. Jha, Julio Frenk, Suerie Moon, Joy Phumaphi, Peter Piot, Barbara Stocking, Victor J. Dzau, and Gabriel M. Leung. Toward a common secure future: Four global commissions in the wake of ebola. *PLOS Medicine*, 13(5):1–15, 05 2016. 72

[36] A. Goubar, D. Bitar, W. C. Cao, D. Feng, L. Q. Fang, and J. C. Desenclos. An approach to estimate the number of SARS cases imported by international air travel. *Epidemiology & Infection*, 137(7):1019–1031, July 2009. 41

[37] E. Gourdin, J. Omic, and P. Van Mieghem. Optimization of network protection against virus spread. *8th International Workshop on Design of Reliable Communication Networks (DRCN 2011), October 10-12, Krakow, Poland.*, 2011. 5

[38] B. T. Grenfell, O. N. Bjørnstad, and J. Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, December 2001. 3

[39] Laurent Hébert-Dufresne, Antoine Allard, Jean-Gabriel Young, and Louis J. Dubé. Global efficiency of local immunization on complex networks. *Scientific Reports*, 3, July 2013. 4, 13, 18, 34

[40] T. Déirdre Hollingsworth, Neil M. Ferguson, and Roy M. Anderson. Will travel restrictions control the international spread of pandemic influenza? *Nature Medicine*, 12(5):497–499, May 2006. 41

[41] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42):15124–15129, October 2004. 2, 3, 27, 52

[42] Swami Iyer, Timothy Killingback, Bala Sundaram, and Zhen Wang. Attack robustness and centrality of complex networks. *PLOS ONE*, 8(4):1–17, 04 2013. 13

[43] Matthew O Jackson and Yves Zenou. Games on networks. 2014. 19

[44] Ayodele Samuel Jegede. What led to the nigerian boycott of the polio vaccination campaign? *PLoS medicine*, 4(3):e73, 2007. 107

[45] M. J. Keeling and C. A. Gilligan. Bubonic plague: a metapopulation model of a zoonosis. *Proceedings of the Royal Society of London B: Biological Sciences*, 267(1458):2219–2230, November 2000. 3

[46] M. J. Keeling and B. T. Grenfell. Disease Extinction and Community Size: Modeling the Persistence of Measles. *Science*, 275(5296):65–67, January 1997. 3

[47] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011. 17

[48] Claudia Keser and Frans Van Winden. Conditional cooperation and voluntary contributions to public goods. *The Scandinavian Journal of Economics*, 102(1):23–39, 2000. 19

[49] Kamran Khan, Rose Eckhardt, John S. Brownstein, Raza Naqvi, Wei Hu, David Kossowsky, David Scales, Julien Arino, Michael MacDonald, Jun Wang, Jennifer Sears, and Martin S. Cetron. Entry and exit screening of airline travellers during the A(H1n1) 2009 pandemic: a retrospective evaluation. *Bulletin of the World Health Organization*, 91(5):368–376, May 2013. 41

[50] Maria A. Kiskowski. A Three-Scale Network Model for the Early Growth Dynamics of 2014 West Africa Ebola Epidemic. *PLoS Currents*, 6, November 2014. 3

[51] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernan A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, November 2010. arXiv: 1001.5285. 4, 13, 18, 34

[52] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Phys. Rev. Lett.*, 87:198701, Oct 2001. 14

[53] Vito Latora and Massimo Marchiori. Is the boston subway a small-world network? *Physica A: Statistical Mechanics and its Applications*, 314(1):109 – 113, 2002. Horizons in Complex Systems. 14

[54] John O Ledyard. Public goods: A survey of experimental research. 1994. 19

[55] Philippe Lemey, Andrew Rambaut, Trevor Bedford, Nuno Faria, Filip Bielejec, Guy Baele, Colin A. Russell, Derek J. Smith, Oliver G. Pybus, Dirk Brockmann, and Marc A. Suchard. Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3n2. *PLOS Pathogens*, 10(2):e1003932, February 2014. 52

[56] High level Panel on the Global Response to Health Crises. Protecting humanity from future health crises. resreport, United Nations, January 2016. 72

[57] David K Levine. Modeling altruism and spitefulness in experiments. *Review of economic dynamics*, 1(3):593–622, 1998. 19

[58] OAG Aviation Worldwide Ltd. The largest airline schedules and flight status database in the world. 27, 28, 41, 103

[59] John D. Malone, Robert Brigantic, George A. Muller, Ashok Gadgil, Woody Delp, Benjamin H. McMahon, Russell Lee, Jim Kulesz, and F. Matthew Mihelic. U.S. airport entry screening in response to pandemic influenza: Modeling and analysis. *Travel Medicine and Infectious Disease*, 7(4):181–191, July 2009. 41

[60] Sandip Mandal, Ram Rup Sarkar, and Somdatta Sinha. Mathematical models of malaria - a review. *Malaria Journal*, 10:202, July 2011. 3

[61] Alexandra Mangili and Mark A Gendreau. Transmission of infectious diseases during commercial air travel. *The Lancet*, 365(9463):989–996, March 2005. 2, 50

[62] Naoki Masuda. Immunization of networks with community structure. *New Journal of Physics*, 11(12):123018, 2009. 13, 18

[63] Naoki Masuda, Mason A. Porter, and Renaud Lambiotte. Random walks and diffusion on networks. *arXiv:1612.03281 [cond-mat, physics:physics]*, December 2016. arXiv: 1612.03281. 4, 96

[64] Martial L. Ndeffo Mbah and Christopher A. Gilligan. Resource Allocation for Epidemic Control in Metapopulations. *PLOS ONE*, 6(9):e24577, September 2011. 5, 23, 72, 87, 105

[65] Martial L Ndeffo Mbah, Jingzhou Liu, Chris T Bauch, Yonas I Tekel, Jan Medlock, Lauren Ancel Meyers, and Alison P Galvani. The impact of imitation on vaccination behavior in social contact networks. *PLoS computational biology*, 8(4):e1002469, 2012. 19

[66] Stefano Merler, Marco Ajelli, Laura Fumanelli, Marcelo F C Gomes, Ana Pastore y Piontti, Luca Rossi, Dennis L Chao, Ira M Longini, M Elizabeth Halloran, and Alessandro Vespignani. Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. *The Lancet Infectious Diseases*, 15(2):204–211, February 2015. 3, 18

[67] Lauren Ancel Meyers, Babak Pourbohloul, M. E. J. Newman, Danuta M. Skowronski, and Robert C. Brunham. Network theory and SARS: predicting outbreak diversity. *Journal of Theoretical Biology*, 232(1):71–81, January 2005. 3, 27

[68] Suerie Moon, Devi Sridhar, Muhammad A Pate, Ashish K Jha, Chelsea Clinton, Sophie Delaunay, Valnora Edwin, Mosoka Fallah, David P Fidler, Laurie Garrett, et al. Will ebola change the game? ten essential reforms before the next pandemic. the report of the harvard-lshtm independent panel on the global response to ebola. *The Lancet*, 386(10009):2204–2221, 2015. 72

[69] M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128, July 2002. 96

[70] Mark Newman. *Networks: An Introduction*. Oxford University Press, Oxford, New York, March 2010. 14, 15, 34, 53, 55

[71] Hiroshi Nishiura and Kazuko Kamiya. Fever screening during the influenza (H1n1-2009) pandemic at Narita International Airport, Japan. *BMC Infectious Diseases*, 11:111, May 2011. 41

[72] Martin A Nowak. *Evolutionary dynamics*. Harvard University Press, 2006. 19

[73] Theo Offerman. *Beliefs and decision rules in public good games: Theory and experiments*. Springer Science & Business Media, 2013. 6

[74] Sonja J. Olsen, Hsiao-Ling Chang, Terence Yung-Yan Cheung, Antony Fai-Yu Tang, Tamara L. Fisk, Steven Peng-Lim Ooi, Hung-Wei Kuo, Donald Dah-Shyong Jiang, Kow-Tong Chen, Jim Lando, Kwo-Hsiung Hsu, Tzay-Jinn Chen, and Scott F. Dowell. Transmission of the Severe Acute Respiratory Syndrome on Aircraft. *New England Journal of Medicine*, 349(25):2416–2422, December 2003. 50

[75] Committee on the Functioning of the International Health Regulations (2005) and on Pandemic Influenza A (H1N1) 2009. Report of the review committee on the functioning of the international health regulations (2005) and on pandemic influenza a (h1n1) 2009. Technical report, WHO, March 2011. 1, 72

[76] United Nations High-Level Panel on the Global Response to Health Crises. *The Neglected Dimension of Global Security: A Framework to Counter Infectious Disease Crises*. The National Academies Press, Washington, DC, 2016. 72

[77] Ebola Interim Assessment Panel. Report of the ebola interim assessment panel. http://www.who.int/csr/resources/publications/ebola/ebola-panel-report/en/, July 2015. Acessed:2018-01-03. 1

[78] Ebola Interim Assessment Panel. Report of the ebola interim assessment panel. Technical report, WHO, July 2015. 72

[79] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925–979, August 2015. 18, 96

[80] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6):066117, May 2001. 96

[81] Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of complex networks. *Phys. Rev. E*, 65:036104, Feb 2002. 18

[82] Reena Pattani. Unsanctioned travel restrictions related to ebola unravel the global social contract. *Canadian Medical Association Journal*, 187(3):166–167, 2015. 72

[83] Ana Perisic and Chris T Bauch. Social contact networks and disease eradicability under voluntary vaccination. *PLoS computational biology*, 5(2):e1000280, 2009. 5, 19

[84] Chiara Poletto, Marcelo F. C. Gomes, Ana Pastore y Piontti, Luca Rossi, Livio Bioglio, Dennis L. Chao, Ira M. Longini, M. Elizabeth Halloran, Vittoria Colizza, and Alessandro Vespignani. Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 19(42), October 2014. 3, 18

[85] B. Prakash, L. Adamic, T. Iwashyna, H. Tong, and C. Faloutsos. Fractional Immunization in Networks. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, Proceedings, pages 659–667. Society for Industrial and Applied Mathematics, May 2013. DOI: 10.1137/1.9781611972832.73. 5, 24

[86] V. M. Preciado and M. Zargham. Traffic optimization to control epidemic outbreaks in metapopulation models. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 847–850, December 2013. 5, 18

[87] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas. Optimal Resource Allocation for Network Protection Against Spreading Processes. *IEEE Transactions on Control of Network Systems*, 1(1):99–108, March 2014. 5, 24, 72

[88] David G Rand and Martin A Nowak. The evolution of anti-social punishment in optional public goods games. *Nature communications*, 2:434, 2011. 72

[89] Jonathan M. Read, Peter J. Diggle, James Chirombo, Tom Solomon, and Matthew Baylis. Effectiveness of screening for Ebola at airports. *The Lancet*, 385(9962):23–24, January 2015. 41

[90] Wendy Rhymer and Rick Speare. CountriesâĂŹ response to whoâĂŹs travel recommendations during the 2013–2016 ebola outbreak. *Bulletin of the World Health Organization*, 95(1):10, 2017. 72

[91] Caitlin M. Rivers, Eric T. Lofgren, Madhav Marathe, Stephen Eubank, and Bryan L. Lewis. Modeling the Impact of Interventions on an Epidemic of Ebola in Sierra Leone and Liberia. *PLoS Currents*, 6, November 2014. 3, 18

[92] Yannick Rochat. Closeness centrality extended to unconnected graphs: The harmonic centrality index. In *ASNA*, number EPFL-CONF-200525, 2009. 13

[93] Robert E. Rowthorn, Ramanan Laxminarayan, and Christopher A. Gilligan. Optimal control of epidemics in metapopulations. *Journal of The Royal Society Interface*, 6(41):1135–1144, December 2009. 5, 23, 72, 87, 105

[94] Robert M. May Roy M. Anderson. *Infectious Diseases of Humans*. Oxford University Press, 1991. 2

[95] Shigui Ruan, Wendi Wang, and Simon A. Levin. The effect of global travel on the spread of sars. *Mathematical biosciences and engineering: MBE*, 3(1):205–218, January 2006. 3

[96] Francisco C Santos, Marta D Santos, and Jorge M Pacheco. Social diversity promotes the emergence of cooperation in public goods games. *Nature*, 454(7201):213–216, 2008. 19

[97] P. H. T. Schimit and L. H. A. Monteiro. A vaccination game based on public health actions and personal decisions. *Ecological Modelling*, 222(9):1651–1655, May 2011. 5

[98] Bernhard Sendhoff, Martin Kreutz, and Werner Von Seelen. A condition for the genotype-phenotype mapping: Causality. *arXiv preprint adap-org/9711001*, 1997. 22

[99] Bita Shams and Mohammad Khansari. Using network properties to evaluate targeted immunization algorithms. *Network Biology*, 2014. 4, 34

[100] John Snow. *On the mode of communication of cholera*. John Churchill, 1855. 17

[101] A. J. Tatem, D. J. Rogers, and S. I. Hay. Global Transport Networks and Infectious Disease Spread. *Advances in Parasitology*, 62:293–343, January 2006. 2, 27, 51

[102] Michele Tizzoni, Paolo Bajardi, Chiara Poletto, José J Ramasco, Duygu Balcan, Bruno Gonçalves, Nicola Perra, Vittoria Colizza, and Alessandro Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 a/h1n1pdm. *BMC medicine*, 10(1):165, 2012. 18

[103] Shouqiang Wang, Francis de Véricourt, and Peng Sun. Decentralized resource allocation to control an epidemic: A game theoretic approach. *Mathematical biosciences*, 222(1):1–12, 2009. 5, 6, 24, 75, 83, 104

[104] Duncan J Watts. Networks, dynamics, and the small-world phenomenon. *American Journal of sociology*, 105(2):493–527, 1999. 14

[105] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998. 14

[106] Sheldon J. Watts. *Epidemics and History: Disease, Power and Imperialism*. Yale University Press, 1999. Google-Books-ID: AVX7NInQraoC. 51

[107] Hui-Ming Wei, Xue-Zhi Li, and Maia Martcheva. An epidemic model of a vector-borne disease with direct transmission and time delay. *Journal of Mathematical Analysis and Applications*, 342(2):895–908, June 2008. 3

[108] Eric W. Weisstein. Lattice graph. From MathWorld–A Wolfram Web Resource. 14

[109] WHO. WHO | About IHR. 1, 52, 71

[110] WHO, editor. *Acute respiratory syndrome in Hong Kong Special Administrative Region of China/ Viet Nam*. WHO, March 2003. 51

[111] WHO. Who: Ebola response roadmap situation report 1. http://www.who.int/csr/disease/ebola/situation-reports/archive/en/, August 2014. Acessed:2018-01-03. 1

[112] WHO. Disease outbreak news: Yellow fever in angola. http://who.int/csr/don/12-february-2016-yellow-fever-angola/en/, February 2016. Accessd:2018-01-04. 2

[113] WHO. Situation report: Ebola virus disease. http://www.who.int/csr/disease/ebola/situation-reports/archive/en/, June 2016. Acessed: 2018-01-03. 1

[114] WHO. WHO | Ebola virus disease, June 2016. 1, 72

[115] Jin Y Yen. Finding the k shortest loopless paths in a network. *management Science*, 17(11):712–716, 1971. 16

[116] Gregory S. Zaric and Margaret L. Brandeau. Resource allocation for epidemic control over short time horizons. *Mathematical Biosciences*, 171(1):33 – 58, 2001. 23

[117] Tianrui Zhang, Lu-Xing Yang, Xiaofan Yang, Yingbo Wu, and Yuan Yan Tang. Dynamic malware containment under an epidemic model with alert. *Physica A: Statistical Mechanics and its Applications*, 470:249–260, 2017. 5

[118] Yao Zhang and B Aditya Prakash. Scalable vaccine distribution in large graphs given uncertain data. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1719–1728. ACM, 2014. 24